



**UCL**

# Naturalistic Depth Perception and Binocular Vision

GUIDO MAIELLO

THESIS SUBMITTED FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

INSTITUTE OF OPHTHALMOLOGY  
UNIVERSITY COLLEGE LONDON

2017

I, Guido Maiello confirm that the work presented in this thesis is my own.  
Where information has been derived from other sources, I confirm that this  
has been indicated in the thesis.

A handwritten signature in black ink, appearing to read 'G. Maiello', written in a cursive style.



# Abstract

Humans continuously move both their eyes to redirect their foveae to objects at new depths. To correctly execute these complex combinations of saccades, vergence eye movements and accommodation changes, the visual system makes use of multiple sources of depth information, including binocular disparity and defocus. Furthermore, during development, both fine-tuning of oculomotor control as well as correct eye growth are likely driven by complex interactions between eye movements, accommodation, and the distributions of defocus and depth information across the retina. I have employed photographs of natural scenes taken with a commercial plenoptic camera to examine depth perception while varying perspective, blur and binocular disparity. Using a gaze contingent display with these natural images, I have shown that disparity and peripheral blur interact to modify eye movements and facilitate binocular fusion. By decoupling visual feedback for each eye, I have found it possible to induce both conjugate and disconjugate changes in saccadic adaptation, which helps us understand to what degree the eyes can be individually controlled. To understand the aetiology of myopia, I have developed geometric models of emmetropic and myopic eye shape, from which I have derived psychophysically testable predictions about visual function. I

have then tested the myopic against the emmetropic visual system and have found that some aspects of visual function decrease in the periphery at a faster rate in best-corrected myopic observers than in emmetropes. To study the effects of different depth cues on visual development, I have investigated accommodation response and sensitivity to blur in normal and myopic subjects. This body of work furthers our understanding of oculomotor control and 3D perception, has applied implications regarding discomfort in the use of virtual reality, and provides clinically relevant insights regarding the development of refractive error and potential approaches to prevent incorrect emmetropization.

# Contents

<b>Abstract</b>	<b>2</b>
<b>Table of Contents</b>	<b>5</b>
<b>Figures and Tables</b>	<b>10</b>
<b>Acknowledgements</b>	<b>17</b>
<b>Introduction</b>	<b>21</b>
<b>1 The Visual System</b>	<b>22</b>
1.1 The Eye . . . . .	23
1.2 Visual Pathways . . . . .	25
1.3 Visual Cortex . . . . .	27
1.4 Binocular Vision . . . . .	29
1.5 The Oculomotor System . . . . .	31
<b>2 Motivation and Overview of Thesis</b>	<b>33</b>

2.1	Motivation . . . . .	33
2.2	Overview . . . . .	36
<b>Part 1: Depth and Blur</b>		<b>44</b>
<b>3</b>	<b>The (In)Effectiveness of Simulated Blur for Depth Perception in Naturalistic Images</b>	<b>45</b>
3.1	Introduction . . . . .	45
3.2	Methods . . . . .	48
3.3	Results . . . . .	60
3.4	Discussion . . . . .	63
<b>4</b>	<b>Simulated Disparity and Peripheral Blur Interact During Binocular Fusion</b>	<b>69</b>
4.1	Introduction . . . . .	69
4.2	Methods . . . . .	71
4.3	Results . . . . .	80
4.4	Discussion . . . . .	87
<b>Part 2: Myopia</b>		<b>93</b>
<b>5</b>	<b>Retinal Blur from Natural Scenes and Eye Shape</b>	<b>94</b>
5.1	Introduction . . . . .	94
5.2	Methods . . . . .	97
5.3	Results . . . . .	104
5.4	Discussion . . . . .	109

<b>6</b>	<b>Blur Perception Throughout the Visual Field in Myopia and Emmetropia</b>	<b>114</b>
6.1	Introduction . . . . .	114
6.2	Methods . . . . .	117
6.3	Results . . . . .	124
6.4	Discussion . . . . .	129
<b>7</b>	<b>Vergence Driven Accommodation with Simulated Disparity in Myopia and Emmetropia</b>	<b>132</b>
7.1	Introduction . . . . .	132
7.2	Methods . . . . .	137
7.3	Results . . . . .	148
7.4	Discussion . . . . .	157
	<b>Part 3: Perception, Action and Motion</b>	<b>161</b>
<b>8</b>	<b>Monocular and Binocular Contributions to Oculomotor Plasticity</b>	<b>162</b>
8.1	Introduction . . . . .	162
8.2	Methods . . . . .	165
8.3	Results . . . . .	171
8.4	Discussion . . . . .	176
<b>9</b>	<b>Three Dimensional Binocular Eye and Hand Coordination in Normal Vision and with Simulated Visual Impairments</b>	<b>180</b>
9.1	Introduction . . . . .	180

9.2	Methods . . . . .	185
9.3	Results . . . . .	191
9.4	Discussion . . . . .	196
<b>10</b>	<b>A Space-variant Model for Motion Interpretation across the Visual Field</b>	<b>202</b>
10.1	Introduction . . . . .	202
10.2	Methods . . . . .	229
10.3	Results . . . . .	237
10.4	Discussion . . . . .	248
<b>Part 4:</b>	<b>Assessment of Low-Cost Technologies</b>	<b>256</b>
<b>11</b>	<b>Evaluation of the Tobii EyeX Eye Tracking Controller and Matlab Toolkit for Research</b>	<b>257</b>
11.1	Introduction . . . . .	257
11.2	Methods . . . . .	261
11.3	Results . . . . .	268
11.4	Discussion . . . . .	275
<b>12</b>	<b>The Perceptual Quality of the Oculus Rift for Immersive Virtual Reality</b>	<b>280</b>
12.1	Introduction . . . . .	280
12.2	Methods . . . . .	286
12.3	Results . . . . .	299
12.4	Discussion . . . . .	319

<b>Final Conclusions and Future Work</b>	<b>324</b>
13 Summary of Results and Conclusions	325
14 Limitations	329
15 Suggestions for Future Research	339
<b>Bibliography</b>	<b>348</b>
<b>Appendix</b>	<b>410</b>
List of Conference Posters and Presentations . . . . .	412
List of Peer-Reviewed Publications . . . . .	415

# Figures and Tables



# List of Figures

1.1	Diagram of the human eye . . . . .	24
1.2	Diagram of the human visual pathways . . . . .	26
3.1	Lytro <sup>TM</sup> plenoptic camera . . . . .	50
3.2	Steps to determine the in-focus area for the first image in the stack . . . . .	51
3.3	Example of extracted in-focus patches for each focal plane . . . . .	52
3.4	Blur in a conventional lens system . . . . .	53
3.5	Dioptric depth separation between image planes of light field pictures . . . . .	55
3.6	Example of one trial from the temporal 2AFC exper- iment . . . . .	58
3.7	Example of one trial from the spatial 2AFC experiment	60
3.8	Depth discrimination sensitivity . . . . .	61
4.1	Focus stacking . . . . .	76
4.2	Geometry of disparity rendering . . . . .	77
4.3	Example of a binocular fusion experimental trial . . . .	79

4.4	Time to perceptual fusion . . . . .	81
4.5	Fixations necessary for perceptual fusion . . . . .	83
4.6	Fixation duration while viewing the stimuli as a function of disparity . . . . .	85
4.7	Vergence stability . . . . .	86
4.8	Horizontal vergence stability of the first fixation of every trial as a function of disparity . . . . .	88
5.1	Model of retinal surface . . . . .	98
5.2	Example range image . . . . .	99
5.3	Steps to geometrically estimate retinal blur . . . . .	101
5.4	Blur distributions throughout the visual field . . . . .	106
5.5	Blur in emmetropic and myopic model eyes . . . . .	108
5.6	Peripheral visual function in myopes and emmetropes . . . . .	110
6.1	Example stimuli for blur discrimination . . . . .	120
6.2	Example of a dipper function . . . . .	122
6.3	Average blur dipper functions . . . . .	125
6.4	Intrinsic Blur . . . . .	126
6.5	Blur Sensitivity . . . . .	128
7.1	Schematization of the experimental setup to measure vergence and accommodation . . . . .	139
7.2	Stimuli to drive vergence and accommodation . . . . .	141
7.3	Accommodation as a function of time for a representative observer in which Step 2 was a “Zero gradient” stimulus . . . . .	147

7.4	Vergence response curves for Steps 1 and 2 . . . . .	149
7.5	Accommodation response curves for Steps 1 and 2 . .	150
7.6	Accommodation and accommodation variability during Step 1 . . . . .	152
7.7	Exponential half-life $\tau$ . . . . .	154
7.8	Accommodation steady state variability . . . . .	156
8.1	Saccade adaptation paradigm . . . . .	166
8.2	Saccade adaptation . . . . .	172
8.3	Independent vergence and saccade adaptation . . . . .	175
9.1	Refractive error distributions . . . . .	184
9.2	Example stimuli for target tracking . . . . .	187
9.3	Schematics of the four experimental target tracking sessions . . . . .	189
9.4	Example raw accuracy data and hinged line fitting . .	192
9.5	Tracking accuracy, blur and refraction . . . . .	194
10.1	The neural space-variant model . . . . .	204
10.2	. . . . .	211
10.2	The retino-cortical mapping . . . . .	212
10.3	Filtering in the cortical domain . . . . .	218
10.4	Optic flow in the cortical domain . . . . .	221
10.5	Deformation subspaces . . . . .	222
10.6	Local cardinal deformations of the cortical optic flow .	224
10.7	Dead leaf stimuli . . . . .	231

10.8 Schematic of a single trial from equivalent noise experiment . . . . .	236
10.9 . . . . .	241
10.9 FRM estimation in human observers and proposed model . . . . .	242
10.10Equivalent Noise Analysis in human observers . . . . .	243
10.11Equivalent Noise Analysis on the model responses . . . . .	245
10.12Sampling Efficiency depends on the portion of the model's cortical area dedicated to the model's fovea . . . . .	247
10.13Foveocentric bias of the neural model as a function of external noise . . . . .	249
11.1 Fixation locations selected to measure accuracy and precision . . . . .	262
11.2 Accuracy and precision of the Tobii EyeX . . . . .	270
11.3 Stereotypical saccade dynamics . . . . .	273
11.4 Stereotypical pursuit traces . . . . .	274
11.5 Stereotypical vergence dynamics . . . . .	276
12.1 Virtual scenarios . . . . .	291
12.2 Heart rate data from Experiment 1 . . . . .	300
12.3 Immersivity Questionnaire data from Experiment 1 . . . . .	301
12.4 Correlation between observers' fear of height and vertigo in Experiment 1 . . . . .	303
12.5 Accelerometer data from Experiment 2 . . . . .	306

12.6	Head movements as a function of time from obstacle	
	onset . . . . .	308
12.7	Head movements for virtual obstacles . . . . .	309
12.8	Heart rate in Experiment 3 . . . . .	312
12.9	Head movements in Experiment 3 . . . . .	314
12.10	Immersivity questionnaire in Experiment 3 . . . . .	316
12.11	Simulator Sickness Questionnaire scores from Experi-	
	ment 3 . . . . .	318

# List of Tables

10.1	Model parameters for Equivalent Noise experiment . .	238
11.1	Commercial eye tracker comparison . . . . .	277
12.1	Multiple Linear Regression assessing the relationship between Heart Rate and Questionnaire Scores . . . . .	305
12.2	Simulator Sickness Questionnaire scores from Experi- ment 2 . . . . .	311

# Acknowledgements

It feels very bizarre that this chapter of my life is wrapping up. These past four years have been amazing, partly because I have immensely enjoyed the work I have been doing, but mostly because of the awesome people that have populated my life. These acknowledgments are about how these people have enriched my life while incidentally I was also doing science.

Special thanks go to Steven for agreeing to supervise me through UCL and across three continents, to Mara for guiding and supporting me towards this opportunity, and to Fuensanta, Lenna and Kristen, without whom much of the work in this thesis would not have happened. Fabio, Manuela and Silvio, I owe you my gratitude for having started me along this path. You initiated my spark of curiosity towards vision and neuroscience, and have greatly shaped my interests and passions. Thank you for being such supportive mentors.

Agostino has been a source of inspiration for the kind of scientist I aim to be. His passion for his work is so strong and obvious; it has made me push myself in my own work. Also, he is a pretty hilarious guy, so that's always a plus. Grande Agus!

The Bex Lab has felt like large and welcoming, and slightly dysfunctional, family.

Pete Bex says of himself:

“Bex is awesome and there’s no denying it.”

Also, with some help from Google translate, he has been known to announce:

“Lo studente è nulla senza l’insegnante. Tutti acclamano il possente Bex!”

What a self-aggrandizing asshat... However, he is not necessarily wrong.

Learning how to do science under Pete’s mentorship has been one of the most fun and engaging experiences. Pete provided me with the means and guidance to follow my own ideas, satisfy my own curiosities, and most of all simply goof off in the lab. If I end up being a decent scientist it will be largely his fault. Thanks Boss.

MiYoung, is one of my favourite people of all time. Her example has taught me how to be a more focused and rigorous scientist, and how to be constantly amazed at the world. Grazie MiYounga!

William, you are an asshat, not unlike Pete. Nevertheless, I have learned an incredible amount from you, and your passion and love of science and critical thinking have been contagious. Also, you and Sasha have become family. So thanks for all that.

Emily has been the best academic older sister. I have looked up to her and her example has set the bar I have aimed to reach as a graduate student. Thank you Em.

Tom and Jenni, it felt like you guys adopted me when I first moved to Boston. Thank you for being awesome.

Michael and Tobias are almost certainly the best computer scientists I



know. You guys are the standard of excellence I strive for.

Luis, thank you for being my dirty uncle. I will sorely miss the shenanigans we always seem to get ourselves into.

Dylan and Tiffany, take care of the boss man on my behalf. Remember to take him to the pub at least once a week or he gets grumpy, and to make fun of his age/forgetfulness/tendency to mumble/programming skills/etc. every so often or he becomes complacent.

Living across the ocean has been amazing, in no small measure because of the support of family and friends from back home.

My mother is simply the best. She is the strongest, toughest, and kindest person I know. Because of her example, I know that I will be able to face whatever happens in life, and I will always be fine. Grazie Mami.

Paola, thank you for believing in us. You are smart, beautiful, and talented, and I almost certainly do not deserve you. My life is more beautiful and happy because of you. Ti amo Paola.

Zio Mauro, zia Anna e cuginetta Giulia, vi voglio un mare di bene. Tornare a casa da voi mi fa sempre stare bene.

Cugino Lorenzo, thank you for the company and camaraderie in a foreign land, and for being exceptional family. There are few people with whom I share so closely my view of the world and my passion for mayhem.

Alex, Luca, Clara, Matia, Giuliano. You guys are some of my oldest friends, and I do not see you often enough. Yet the beautiful thing is that whenever I do get to see you, it feels like no time at all has passed.

Lorenzo, you are stubborn as a mule and you are the only person that I ever lose my temper with because you alone have perfected the art of getting

on my nerves. Nevertheless, I could not ask for a better brother. I wish our father could see the two of us now. We turned out to be pretty cool, I think he would have enjoyed it. I am very proud of you, brother.

As always, in everything I try to accomplish, I strive to keep the memory of my Father.

Boston, 28 November 2016

# **Part 1: Introduction**

# Chapter 1

## The Visual System

Light, a form of electromagnetic radiation, floods our environment. When light encounters physical objects, it interacts with them in a lawful, predictable manner. Light may be transmitted through, absorbed by, or reflected off the objects it encounters, depending on the physical properties of the objects. The intensity and spectrum of the light coming from an object in the world will be determined by the properties of the light source, by the structure of the world, and by the material properties of the object. The human visual system has evolved to detect and measure the light in the environment, thus enabling us to estimate the physical structure of the environment. This chapter provides a schematic description of the human visual system.

## 1.1 The Eye

The human eye, schematized in Figure 1.1, is the organ through which we visually perceive the world. The eye contains both the optical structures to form an image of the world, as well as the receptors and neural substrate that transduce and encode light energy into neural signals. The shape of the eye can be roughly approximated by a spherical ball linked to a smaller curved frontal unit. The shape of the eye is maintained by the *vitreous gel*. The transparent front-most layer of the eye is the *cornea*. Light reaches the *pupil* through the cornea. The iris is a diaphragm of variable diameter which controls the size of the pupil, varying the amount of light entering the eye. After having passed through the pupil, light reaches the *lens*, a transparent, biconvex structure that, along with the cornea, refracts light in order to focus it onto the *retina*.

The detection of light takes place at the retina (Dowling 2007). The retina consists of a light-sensitive layer of tissue on the inner surface of the eye, composed of complex interconnected layers of neurons. Photoreceptor cells are the neurons directly sensitive to light. There are two types of photoreceptors: *rods*, which function in dim light and provide black-and-white vision, and *cones*, which are used for daytime vision and colour perception.

The central part of the retina with the highest resolution is called the *fovea*. The fovea is located in the centre of the *macula*. The fovea consists almost exclusively of cone photoreceptor cells and is responsible for central, high resolution colour vision. Moving away from the fovea and into the peripheral retina, rod photoreceptors detect light without supporting colour sensitivity and at lower spatial resolution, but with a higher sensitivity to

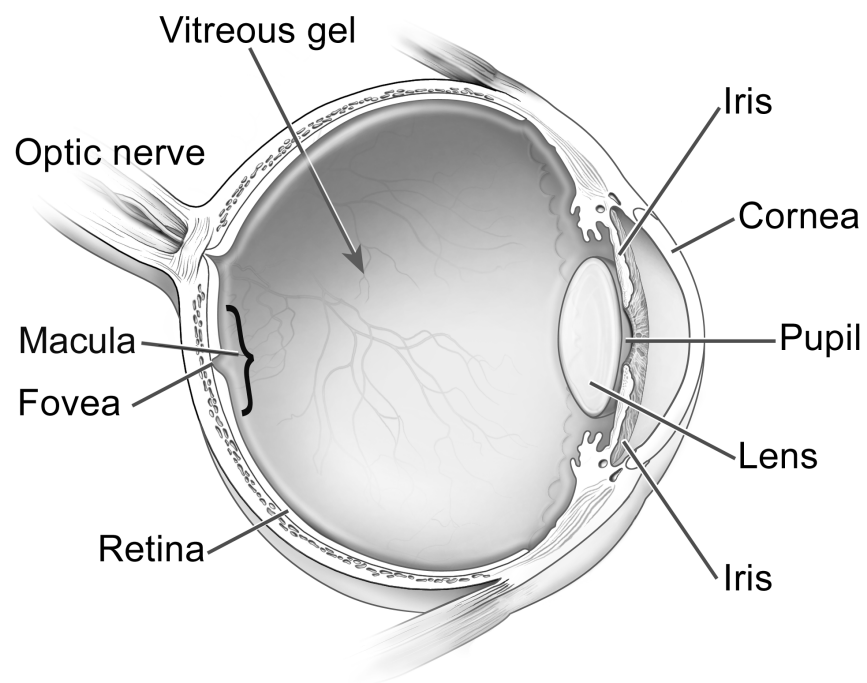


Figure 1.1: **Diagram of the human eye.** Courtesy of the National Eye Institute (NEI).

light. Photoreceptors are located at the very back of the retina. The outer segments of the photoreceptors contain photopigments, which chemically react to light and initiate the transduction of light energy into neural signals. Photoreceptors communicate to bipolar cells, which in turn synapse onto ganglion cells. Lateral interconnections mediated by other neurons, namely horizontal and amacrine cells, are present at each synaptic stage. Ganglion cell axons converge into the optic nerve and transmit the neural signal out of the retina.

## 1.2 Visual Pathways

Humans have two laterally displaced eyes. Thus, each eye has a slightly different view of the world. The field of view of each eye spans, in the vertical direction, 60 degrees superiorly and 70 degrees inferiorly. Along the horizontal direction instead the monocular field of view spans 60 degrees nasally and 100 degrees temporally. An interesting feature of the eye's optics is that image of the world projected onto the retina is flipped. The nasal visual field is projected onto the temporal retina and the temporal visual field is projected onto the nasal retina. The superior retina views the inferior visual field, whereas the inferior retina views the superior visual field.

Figure 1.2 shows a schematization of the visual pathways. The optic nerves from the two eyes meet at the *optic chiasma*. The information coming from the two eyes is rerouted through the optic chiasma according to the visual field. The retinal input corresponding to the left visual field projects to the right brain hemisphere, and the left brain hemisphere receives instead

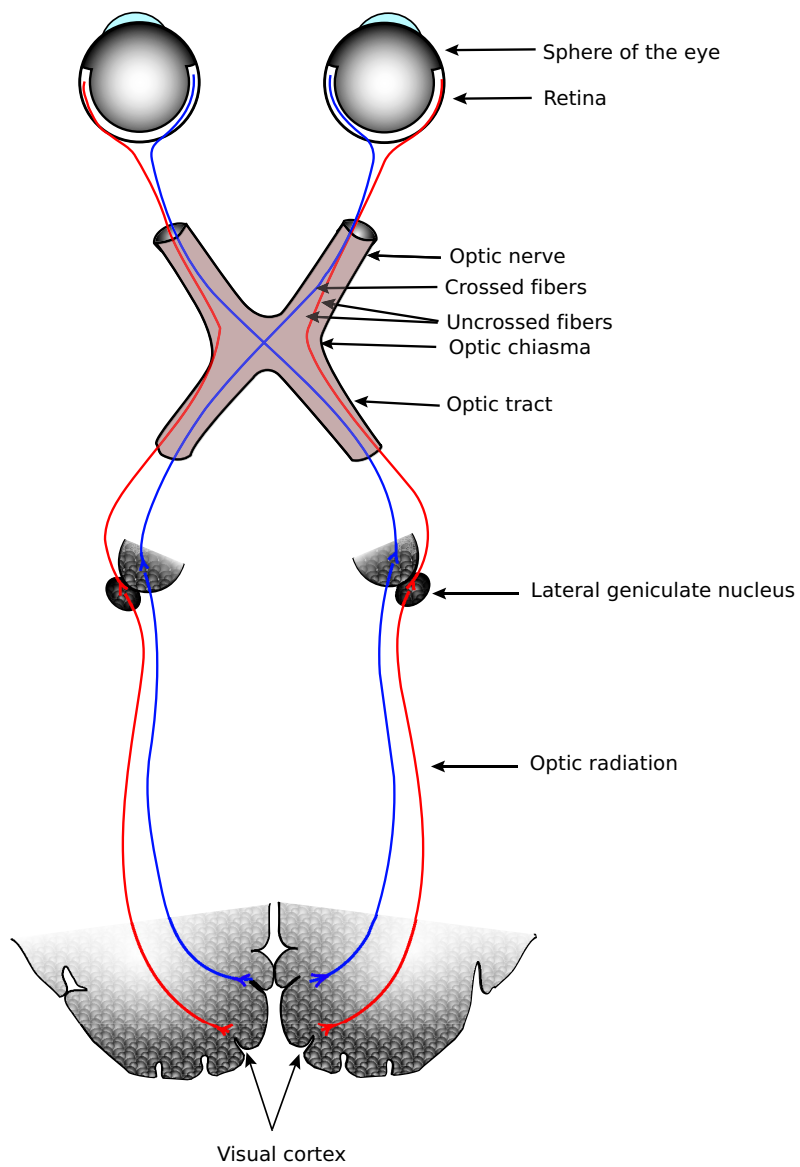


Figure 1.2: **Diagram of the human visual pathways.**



retinal input from the right visual field. To achieve this, optic nerve fibers from the nasal retina in each eye cross over to the opposite side of the brain at the optic chiasma (*crossed fibers*, blue in Figure 1.2). Optic nerve fibers coming from the temporal retina instead project through the optic chiasma to the same side of the brain (*uncrossed fibers*, red in Figure 1.2). Thus, nerve fibers from the right and left visual fields travel into the left and right *optic tracts* respectively. The two optic tracts terminate in the *lateral geniculate nucleus* (LGN). The LGN is a sensory processing and relay center in the thalamus of the brain. The *optic radiations* consists of axons from the neurons in the LGN that carry the visual information onto the visual cortex.

All visual sensory neurons, including photoreceptors, retinal ganglion cells, neurons throughout the visual pathways and neurons in visual cortical areas, alter their firing rate selectively to visual stimuli within a region of retinal space, and thus a small area of visual space that depends on eye posture. This region of visual space is called the receptive field of a visual neuron.

### 1.3 Visual Cortex

Visual cortex, located in the occipital lobe at the back of the skull, is the part of the cerebral cortex that processes visual information.

Primary visual cortex, known as V1, is the first cortical region involved in processing visual sensory input coming directly from the thalamus (Carandini 2012). V1 has a retinotopic organization, meaning that all retinal locations map onto specific portions of V1, and this retinotopy is continuous. However,

this mapping is distorted: a small portion of the central, foveal visual field is mapped onto a large portion of V1, whereas increasingly fewer V1 neural resources are allotted to farther eccentricities in the peripheral visual field. The specific nature and significance of this distortion is explored in Chapter 10. V1 neurons are typically classified as either simple or complex cells. V1 simple cells respond primarily to edges and gratings and have clear inhibitory and excitatory regions within their receptive fields. Complex cells combine the input from multiple simple cells and thus also respond to edges and gratings, but their response is position invariant, i.e. the cell will respond to visual stimulus no matter where it is presented within its receptive field. The primary role of V1 is to perform a first low level analysis of the visual signal from the retina. Thus V1 neurons are selective for a number of low level image attributes including orientation, spatial and temporal frequency, direction of motion, binocular disparity and color. V1 can be thought of as performing basic filtering operations to extract low level image features, which are then sent to higher cortical areas.

According to the ventral-dorsal model, visual information from V1 is then sent along to two main processing streams (Ungerleider and Pessoa 2008). The ventral stream, also called the *What Pathway* is broadly associated with object recognition. Information along the ventral stream broadly goes through areas V2 and V4 (that are roughly specialized in processing form and colour) and inferior temporal (IT) cortex. The dorsal stream, also known as the *Where Pathway* is instead associated with spatial aspect of visual stimuli such as location and motion. Information along the dorsal stream goes from V1 onto visual cortical areas V2, V3, V3a, middle temporal area

(MT/V5) and medial superior temporal area (MST), all of which show selectivity for motion and depth. The role of areas MT and MST for processing optic flow information is explored in Chapter 10. Although the two stream hypothesis is a useful model with which to explore the organization of the visual system, it is a obvious oversimplification, as the two processing streams are interconnected and by no means functionally independent (Schenk and McIntosh 2010).

## 1.4 Binocular Vision

In vertebrates, periscopy, the ability to sample the world at 360 degrees, is often sacrificed in favour of some degree of binocularity and stereoscopy. Humans, as already mentioned, have two front-facing eyes with overlapping visual fields. This occurs because binocular vision confers multiple advantages. Stereoscopy can in fact aid to break camouflage without the need to move one's viewpoint (Julesz 1971). Furthermore, combining information from the two eyes reduces noise and improves sensitivity by a factor of  $\sqrt{2}$  (Legge 1984a, 1984b). Lastly, binocularity provides a signal related to distance. Geometrically, it is possible to reconstruct a lawful relationship between the depth structure of the world and the images on the back of the eyes. Thus, binocularity can be exploited for depth perception. These geometrical relationships between depth and binocularity are described thoroughly in chapters 3 and 4. The key notion however, is that our two eyes are horizontally displaced and thus have two slightly different views of the world. When our two eyes are fixating on a point in space, the image of this

point falls upon corresponding locations on the center of the retina (fovea) in each eye. The images cast by nearer or farther points instead will not fall on corresponding points in the two retinas. The distance between the projection of world points on the two retinæ is called binocular disparity, and from this quantity the position in depth of all objects in the world can in principle be calculated.

The computation of binocular disparity however suffers from the correspondence problem: to estimate disparity, the visual system must determine which points on the two retinæ match up in the real world, in conditions in which multiple matches are possible. One possible method of solving the correspondence problem would be to first perform object recognition and then match up objects in the two eyes. However, the fact that we can perceive depth in random dot stereograms (Julesz 1971), which provide no way of extracting objects, implies that the stereo correspondence problem is solved early in the visual pathway, prior to object recognition. A more likely strategy employed by the visual system is that of performing local image correlation at multiple spatial scales between the images on the left and right retinæ: the disparity energy calculations performed by V1 complex cells are in fact equivalent to detecting inter-ocular correlation (Ohzawa, DeAngelis, Freeman, et al. 1990). Disparity could be encoded by V1 simple cells either by combining receptive fields with the same shape but shifted in position (position encoding), or by combining receptive fields centered in the same position but with different phase (phase encoding). In reality, many neurons exhibit both phase and position sensitivity (Fleet, Wagner, and Heeger 1996). Simple cells are then combined into complex cells under the disparity energy

model (Ohzawa, DeAngelis, Freeman, et al. 1990). The neural encoding of binocular disparity gives rise, through further processing, to the perception of stereoscopic depth, and is also employed by the oculomotor system to plan eye movements in three dimensional space.

## 1.5 The Oculomotor System

As previously discussed, humans have high resolution central vision and low resolution peripheral vision (Weymouth 1958). To visually explore the environment in detail, humans must therefore move their eyes multiple times per second, rapidly shifting objects of interest selected with the low resolution peripheral retina onto the high resolution fovea (Dodge 1903). Thus, the oculomotor system continuously executes fast, ballistic *saccadic* eye movements (Javal 1879; Kowler 2011). To track moving objects the visual system executes slow, *smooth pursuit* eye movements, which hold the fovea on a tracked object by matching its speed (Lisberger 2015). Furthermore, because humans have binocular overlapping visual fields, to inspect objects at different depths the oculomotor system also executes *vergence* eye movements, which are slow rotations of the two eyes in opposite directions that shift the binocular gaze point in depth (Coubard 2013). Thus humans must continuously execute highly precise and coordinated fast saccadic, slow vergence and slow pursuit eye movements to shift their gaze point in three-dimensional space and keep the image of a visual target focused on the two foveae. Lastly, the oculomotor system executes two additional types of eye movements, vestibulo-ocular reflexes and optokinetic nystagmus, which serve to stabilize the eyes when

the head moves or appears to move.

The oculomotor system that plans, coordinates and executes eye movements is of nontrivial complexity (Schor 2011). Three pairs of extraocular muscles for each eye allow the eyes to rotate around one of three axes, thus providing each eye with 3 degrees of freedom. The motor neurons in the nuclei of cranial nerves III, IV and VI directly innervate and control the six extraocular muscles. These motor neurons are in turn innervated by premotor nuclei in the brainstem which coordinate the combined actions of the extraocular muscles to execute complex horizontal, vertical, and torsional eye rotations. Finally, premotor neurons receive input from multiple distributed brain regions.

In the midbrain, the superior colliculus is involved in the control of all types of eye movements, and the substantia nigra modulates the control of saccadic eye movements via sustained inhibitory input (Sparks 1999; Hikosaka, Takikawa, and Kawagoe 2000). The cerebellum is mainly involved in the online control, rather than the initiation, of eye movements (Robinson and Fuchs 2001). Thus the cerebellum strongly contributes to smooth pursuit eye movements as well as the continuous recalibration of the saccadic system (investigated in Chapter 8). In the frontal cortical regions, the frontal eye fields (FEF) have a direct role in the voluntary generation and control of saccadic eye movements, whereas the supplementary eye fields (SEF) likely have a role in monitoring oculomotor performance (Schall and Thompson 1999). Extrastriate cortical areas involved in eye movement planning include MST, the lateral intraparietal area (LIP), and the posterior parietal area (PP).

# Chapter 2

## Motivation and Overview of Thesis

### 2.1 Motivation

Children and adults spend an increasing proportion of their time viewing information presented on flat screens. With technology allowing nearly 24-hour media access, Rideout, Foehr, and Roberts (2011) report that the amount of time young people in the US spend with entertainment media has risen dramatically. In 2009, 8-18 year-olds devoted on average  $\sim 7.5$  hrs/day using entertainment media, of which  $\sim 4.5$  hrs was spent watching TV, 1.5 hrs using computers,  $\sim 1.25$  hrs on video games, and  $\sim 40$  minutes on print (n. b. 29% of media time was spent using more than one medium concurrently). These trends are likely to continue as access to media increases (e.g. via smart-phones whose users were estimated to number 2.2 billion in 2016).

People's visual experience when using such computer/entertainment de-

vices is very different from natural viewing conditions in which there is a broad and continuously varying range of visual depth information. For example, many artificial images contain little blur because a user often has to be able to extract information from all portions of the image (e.g. when using a desktop application or when playing a videogame). When reading a book or using an e-reader, the pages are also focused all over, and the accommodative distance is much closer to us than in natural, outdoors, conditions (Geisler 2008). In directed movies and television, spatial blur is manipulated to induce the viewer to attend to specific, in-focus portions of the scene. Such conditions differ from the real-world - where the ranges of depths at which our eyes accommodate produce spatio-temporal variation in retinal blur - and essentially de-couple image-blur from accommodation. Furthermore, stereoscopic 3D is increasingly used to promote the illusion of depth, but current display technology nearly always presents depth information uncoupled from focus information. Even the best current 3D displays require the viewer's accommodation be locked onto one focal distance (the position of the monitor) for long periods of time.

There are at least three fundamental differences between indoor visual experience and traditional activity in outdoor environments: the distribution of depths in natural scenes (Huang, Lee, and Mumford 2000) generates:

1. specific changes in accommodation as objects at different depth planes are fixated.
2. a specific range of dioptric blur levels across the peripheral retinae by objects that are closer to or further from the observer than the object currently being fixated.



3. a specific range of depth levels across the peripheral retinae by objects that are closer to or further from the observer than the object currently being fixated.

Visual depth information is a fundamental aspect in the development of the visual system. Binocular coordination in children is less accurate than in adults (Yang and Kapoula 2003; Fioravanti et al. 1995), and the disruption of binocular visual input during development leads to binocular oculomotor dysfunctions (Birch 2013), such as amblyopia, strabismus and vergence insufficiency. Visual depth information is also implicated in the development of myopia. The balance of recent evidence supports the notion that myopia results from a combination of genetic and environmental factors (Morgan, Ohno-Matsui, and Saw 2012). Environmental factors include intensity of education and amount of time spent outdoors, and more broadly, the main environmental risk factor is near work: reading, writing, work with computer monitors and videogame play. Recent studies have identified myopia as a major public health concern, due to its rapidly increasing prevalence over the past half century. In the urban areas of developed countries in east and southeast Asia, 80-90% of children completing high school are now myopic, and up to 20% can have high myopia (myopia requiring -6 diopters or more correction) (Morgan, Ohno-Matsui, and Saw 2012). Between 1971-1972 and 1999-2004 the estimated prevalence of myopia in the US population (in persons aged 12 to 54 years) has grown from 25% to 42% (Vitale, Sperduto, and Ferris 3rd 2009).

In this context, the broader aim of the research presented in this thesis has been to investigate the effects of visual depth information on perceptual

and oculomotor performance. In a set of interconnected yet diverse studies I have investigated how the visual system parses and extracts multiple kinds of information from the retinal input. I show how depth, blur and motion information might all be employed by the visual system to guide visual development, eye growth, motor and particularly oculomotor behaviour in three dimensional space.

## **2.2 Overview**

The experimental section of this thesis is divided into four parts. In Part 1 I investigate and discuss the role of blur with regards to depth perception and binocular oculomotor control. In Part 2 I study the potential links between dioptric blur, the depth structure of the world and myopia. In Part 3 I investigate how the visual system might link depth and motion information to action and oculomotor control. In Part 4 I evaluate and discuss novel low-cost technologies that may be employed in both basic and clinical science regarding binocular vision and oculomotor control.

Each part of the thesis contains two or more chapters. Each chapter is a self-contained study. These experimental chapters (chapters 3 through 12) have been structured as full research papers, with thorough introduction, methods, results and discussion sections. Thus, each chapter may be either read in the context of this thesis or as stand-alone pieces of work.

In Chapter 3 I examine depth perception in images of real scenes with naturalistic variation in pictorial depth cues, simulated dioptric blur and binocular disparity. I employ light field photographs of natural scenes taken

with a plenoptic camera that simultaneously captures images at up to 12 focal planes. With these naturalistic stimuli I show that depth discrimination sensitivity is highest when geometric and stereoscopic disparity cues are both present, and that simulated blur cues impair sensitivity by reducing the contrast of geometric information at high spatial frequencies. I then discuss the implications of these findings for virtual reality rendering technology.

In Chapter 4 I develop and study a gaze-contingent display in which natural images are presented to the observer with dioptric blur and stereoscopic disparity that are dependent on the three-dimensional structure of natural scenes. The system simulates a distribution of retinal blur and depth similar to that experienced in real-world viewing conditions by emmetropic observers. I implement the system using the same light-field photographs from Chapter 3, taken with a plenoptic camera which supports digital refocusing anywhere in the images. I couple this capability with an eye-tracking system and stereoscopic rendering. With this display, I examine how the time course of binocular fusion depends on depth cues from blur and stereoscopic disparity in naturalistic images. I show that disparity and peripheral blur interact to modify eye-movement behavior and facilitate binocular fusion, and the greatest benefit is gained by observers who struggled most to achieve fusion. Even though plenoptic images do not replicate an individual's aberrations, the results demonstrate that a naturalistic distribution of depth-dependent blur may improve 3-D virtual reality, and suggest that interruptions of this pattern which flatten the distribution of retinal blur may adversely affect binocular fusion.

Chapter 5 investigates the distributions of retinal blur experienced by

myopic and emmetropic eyes both indoors and outdoors. Retinal blur occurs because humans continuously refocus their eyes to inspect objects at different depths. Whereas objects fixated with the high-resolution fovea are sharply focused, objects at different depths in the visual periphery are blurred. The amount of defocus blur depends on the depth structure of the visual scene and on the shape of the retina. Myopic eyes are elongated compared to the eyes of emmetropic subjects, which may affect the distribution of retinal blur in real scenes. Using 3D maps of real scenes, I demonstrate that the different topography of outdoor and indoor natural environments produces different patterns of blur across the retinae of emmetropic and myopic model eyes. These differences exist regardless of single-focus, central optical correction. In the real world, the peripheral retina of myopic eyes is more blurred than that of emmetropic eyes, and I confirm this computational prediction with functional measurements of contrast sensitivity. I discuss how these observations have important implications in the development of myopia, as eye growth is driven by the pattern of retinal blur across the retina and time spent outdoors is protective against myopia. I speculate that myopia might be driven by patterns of retinal blur that do not conform to those expected in outdoors natural scenes.

Because I find that myopes and emmetropes experience different patterns of peripheral blur, in Chapter 6 I evaluate the ability of myopic and emmetropic observers to detect and discriminate blur in the central and peripheral retina under monocular and binocular viewing conditions. I measure monocular and binocular blur discrimination thresholds as a function of pedestal blur using dead leaves stimuli and an adaptive 4AFC procedure. The

adaptive procedure efficiently estimates a dipper shaped blur discrimination threshold function with two parameters: Intrinsic Blur and Blur Sensitivity. I find that the amount of Intrinsic Blur increases for retinal eccentricities beyond 4 degrees and is lower in binocular than monocular conditions, but is similar across refractive groups. Blur Sensitivity decreases with retinal eccentricity and is highest for binocular viewing in the fovea. I also find that myopes have worse Blur Sensitivity than emmetropes monocularly but not binocularly. Overall I find that blur perception worsens in the visual periphery because optics, sampling, and neural processing are degraded in the visual periphery, and I discuss how myopes exhibit a monocular impairment in Blur Sensitivity which may be involved in the development of refractive error.

Chapter 7 deals with the implications of the vergence-accommodation mismatch with simulated disparity for myopia. The formation of corresponding and focused foveal images requires a close synergy between the vergence and accommodation systems. This linkage may be dysfunctional in people who are at risk of developing refractive errors, and is usually decoupled in virtual reality systems. I therefore study how refractive error affects vergence-accommodation interactions in stereoscopic displays. I measure vergence and accommodative responses while myopic and emmetropic observers view naturalistic stimuli on a 3D display. Overall, I show that vergence and accommodation responses are disrupted in virtual reality displays in both refractive groups, with a greater disruption in myopes than in emmetropes. Furthermore I discuss how accommodation responses are less stable in myopes, perhaps due to a lower sensitivity to dioptric blur, and how

such inaccuracies of accommodation may cause long-term blur on the retina, which has been associated with a failure of emmetropization.

In Chapter 8 I examine oculomotor plasticity when error signals are independently manipulated in each eye, which can occur naturally owing to aging changes in each eye's orbit and extra-ocular muscles, or in oculomotor dysfunctions. I find that both rapid saccades and slow vergence eye movements are continuously recalibrated independently of one another and corrections can occur in opposite directions in each eye. I discuss how existing models assume a single cortical representation of space employed for the control of both eyes, whereas the findings I present provide evidence for independent monocular and binocular plasticities and dissociable spatial mapping for each eye.

Eye movements are guided not exclusively by visual input; the oculomotor system also receives efference copy from the hand motor system. Thus, the motor systems controlling the eyes and hands are linked when healthy humans track self-generated horizontal hand motion. In Chapter 9 I investigate whether this link exists for 3D eye and hand movements in the sagittal plane, and whether these links are resilient to visual impairment. I examine hand-eye coordination with binocular and monocular simulated visual impairment for horizontal pursuit eye movements and for vergence eye movements in depth. I show that hand motor control signals are employed by the oculomotor system to plan and improve the accuracy of both pursuit and vergence eye movements. Monocular blur affects left-right tracking less than binocular blur, whereas monocular blur impairs tracking in depth as much as binocular blur. I also observe a critical blur level up to which pursuit and

vergence eye movements maintained tracking accuracy independent of blur level. I therefore discuss how hand-eye coordination may be vulnerable to underestimated functional deficits associated with eye disease and may be employed to rehabilitate binocular visual impairments.

The processes driving and recalibrating the performance of the motor, oculomotor and perceptual systems described in previous chapters must also make use of visual motion information. In Chapter 10 I investigate a recent neural model for estimating motion and self motion from optic flow. The model incorporates the deep hierarchy of the first stages of the dorsal visual pathway and is space-variant, since it takes into account the retino-cortical transformation of the primate visual system through log-polar mapping. I assess the model by comparing its performance at estimating the focus of radial motion (FRM) with respect to the precision with which human observers can estimate the FRM in naturalistic motion stimuli. I find that the model's performance at FRM estimation as a function of retinal eccentricity neatly maps onto data from human observers. By employing Equivalent Noise Analysis I observe that loss in FRM accuracy for both model and human observers is attributable to a decrease in the efficiency with which motion information is pooled with increasing retinal eccentricity in the visual field. The decrease in sampling efficiency is thus attributable to receptive field size increases with increasing retinal eccentricity, which are in turn driven by the lossy log-polar mapping that projects the retinal image onto primary visual areas. Disparity processing may be included in the model, thus allowing us to study plausible neural mechanisms for the estimation of motion in depth.

All the work presented in this thesis revolves around binocular vision,

depth perception and action in three dimensional space. The findings I present have implications for basic science, but also for commercial and clinical applications. Thus I am interested in facilitating the dissemination of scientific findings. In chapters 11 and 12 I thus evaluate two commercial low-cost devices which may be employed to translate basic vision science into clinical and commercial applications.

In Chapter 11 I review the characteristics of the Tobii Eyex Controller, a new low-cost binocular eyetracker marketed for integration in gaming and consumer applications. I show that the performance of the EyeX regarding accuracy, precision, latency, and sampling frequency is sufficient for some classes of research application. I find that the device can be successfully employed to measure fixation parameters, saccadic, smooth pursuit and vergence eye movements. Despite its limitations with respect to high-end devices, I discuss how the EyeX has the potential to further the dissemination of eye tracking technology to a broad audience, and could be a valuable asset in consumer and gaming applications as well as a subset of basic and clinical research settings.

Lastly, in Chapter 12 I investigate the Oculus Rift, a high performance yet commercially available head-mounted-display. I address two issues concerning the perceptual quality of the Oculus Rift. (i) Is the Oculus able to generate an high degree of immersivity? In particular, is it possible to elicit the sensation of presence via the virtual stimuli rendered by the device? (ii) Does the Virtual Reality experienced through the Oculus Rift induce physical discomfort? To answer these questions, I employ both custom and established questionnaire techniques, as well as physiological monitoring of



heart rate and user motion. I further compare the Oculus Rift against two other low-cost devices used in immersive Virtual Reality: the Google cardboard and a standard 3DTV monitor. The result I report suggest that users find the Oculus Rift to be immersive and realistic. Observers do not experience simulator sickness when the exposure to virtual reality is short and does not induce excessive amounts of vection. Compared to the other devices the Oculus Rift elicits a greater degree of immersivity. I discuss how these results suggest that the Oculus Rift is a potentially powerful tool for a wide array of basic research and clinical applications.

## Part 1: Depth and Blur

## Chapter 3

# The (In)Effectiveness of Simulated Blur for Depth Perception in Naturalistic Images

### 3.1 Introduction

It has been demonstrated that blur is a reliable cue to depth even near fixation, although its effective usefulness is strongly debated (Marshall et al. 1996; Mather 1996; Mather and Smith 2000; Hoffman and Banks 2010; Nefs 2012; Held, Cooper, and Banks 2012; Vishwanath 2012; Langer and Siciliano 2015). Both Marshall et al. (1996) and Mather (1996) have shown that when other cues to depth are removed, occlusion edge blur (i.e. the amount of blur at the borders between blurred and sharp image regions) can

elicit strong and consistent depth ordering effects. With regards to simulated blur, Marshall et al. (1996) have also shown that some observers are biased to perceive blurred textures as near when they are isolated from context. Mather and Smith (2000) have however shown that the visual system does not appear to integrate depth cues from blur and disparity, primarily employing the disparity cue over the blur cue when both are available. Nefs (2012) has shown that manipulating the depth of field in photographs of semi-natural scenes alters perceived depth/width ratios. Held, Cooper, and Banks (2012) recently employed a volumetric stereo display and found that away from fixation, blur is a more precise cue to depth than disparity, and that when both cues are available, the visual system relies on the more informative cue. In the experiments conducted by Held, Cooper, and Banks (2012) and Hoffman and Banks (2010), blur only ever increased with depth and the target with greater blur was always at a greater depth. However, in real environments, blur increases with distance in either direction away from the accommodation plane. Thus, a more blurred target could either be closer to or further from the observer than the object at his/her fixation plane. Indeed Vishwanath (2012) criticized the Held, Cooper, and Banks study by pointing out that what was effectively measured were blur discrimination thresholds, not perceived depth from blur. Langer and Siciliano (2015) have attempted to replicate the findings from Held, Cooper, and Banks (2012) employing conventional display technology, but found that with simulated blur subjects were unsuccessful at employing blur to discriminate depth at distances far from fixation. It is also worth noting that stereoscopic acuity covaries with spatial frequency (Felton, Richards, and Smith 1972; Schumer

and Julesz 1984; Badcock and Schor 1985), and since blur removes high spatial frequencies, blur could potentially impair depth perception.

Apart from notable exceptions (Wang et al. 2011; Lebreton et al. 2014; Zhang et al. 2014), the experiments performed on blur and disparity have generally used impoverished synthetic stimuli, such as lines, gratings and random dot stereograms. These stimuli intentionally avoid monocular cues such as geometric perspective, which are present in everyday viewing of natural scenes. However, it has been suggested that the perception of visual space is determined by priors based on the probability distribution of real-world sources of retinal images (Howe and Purves 2002). I therefore examined depth cue combination with more complex naturalistic stimuli in which I was able to independently manipulate alternative sources of depth information. In particular, I tested the relative contribution of spatial structure (relative size and geometry) due to perspective, binocular disparity, and defocus blur to depth perception using both a temporal and a spatial two-alternative forced choice (2AFC) paradigm. By comparing depth discrimination with and without both defocus blur and disparity in naturalistic images, I aimed to assess the role of each of these simulated depth cues in naturalistic, yet virtual, viewing conditions.

Understanding the interactions between these cues is particularly important when dealing with computer generated stereoscopic three dimensional scenes. In virtual reality environments for example, observers can experience visual fatigue, nausea and diplopia when depth cues are not accurately simulated (Emoto, Niida, and Okano 2005; D. M. Hoffman et al. 2008; Ukai and Howarth 2008; Bando, Iijima, and Yano 2012). Whereas there is evidence

that simulated naturalistic blur can be employed to facilitate binocular fusion of stereoscopic images (Maiello et al. 2014) (see Chapter 4), here I examine whether rendered blur can successfully be employed to enhance depth perception in virtual reality applications. Disparity, whether it arises from natural viewing or artificially rendered through display technology, is a remarkably robust cue to depth. Blur arising from the optics of an individual’s eye might also be a reliable cue to depth, but replicating an individual’s optical aberrations in display rendering technology is currently feasible only with adaptive optics methods (Sawides et al. 2011). Thus assessing whether depth perception from rendered blur is also robust in naturalistic virtual reality may be useful towards bettering current virtual reality technology.

## **3.2 Methods**

### **Subjects**

Four subjects completed the first experiment, one of whom (GM) was an author of this work. Four subjects completed the second experiment, one was an author (GM) and one had participated in the first experiment. All subjects reported normal or corrected to normal acuity and normal stereo vision.

### **Apparatus**

The experiments were created using Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) with Matlab version R2011a (Mathworks). Stimuli

were presented on a Samsung SynchMaster 2233 LCD monitor with a resolution of 1680x1050 pixels at 120 Hz, run from an NVidia Quadro FX580 graphics processing unit. Observers were seated 50cm in front of the monitor with their heads set in a chin rest. Display dot pitch was 0.282 mm and the monitor subtended 49x33 degrees. Stereoscopic disparity was presented via the NVIDIA 3D Vision kit. The cross talk of the system was 1% measured with a Spectrascan 6500 photometer using the methods described by Weissman and Woods (2011).

## Stimuli

The stimuli for the psychophysical experiments were extracted from light field photographs of natural scenes taken with Lytro plenoptic camera (Lytro Inc, CA). The Lytro is a light field camera that simultaneously captures several versions of the same image, each with a different focus. This property allowed us to vary the level of defocus blur across the image in proportion to the distance in depth from the plane of focus (Figure 3.1).

A software tool was used to extract data and *.jpg* files from the Lytro file format (Patel 2012). For each photograph, a *.jpg* image was generated at each focal plane, a depth map of the scene, a look up table that specifies the depth at which each image in the stack is in focus, and the camera settings (including exposure, aperture, focal length, ISO) for each photograph. The image stack produced by the Lytro contained between 1 and 12 images, depending on the depth structure of the scene. If the scene contained many objects at depths near to the camera, more images were output. For panoramic scenes, where everything is effectively in focus at infinity, only



Figure 3.1: **Lytro™ plenoptic camera.** Each photographic exposure generated a stack of up to 12 images, each in focus at a different depth plane.

one image was generated.

The photographer (author GM) positioned the camera at eye level at locations where the closest object was approximately 10 cm from the lens. Ten light field photographs were selected by the author to include only those in which 11 or 12 focal plane images were generated. For each photograph, the depth map and the depth look up table were used to calculate the location corresponding to the focal depth of each image in the stack. This involved finding the pixel location in which each image of the stack had an in-focus patch. To do this, the first step was resizing the depth map, using bicubic interpolation, from its original size of 20x20 pixels (Figure 3.2a) to be the same size as the focal plane images (1080x1080 pixels). The obtained depth image was further smoothed with a 2D Gaussian low-pass filter of size 3x3 pixels and standard deviation of 0.5 pixels (Figure 3.2b).

For each image in the stack, the corresponding focal depth value was retrieved from the depth look up table. The depth value corresponded to the focal depth at which the image was in-focus. The distance of this depth



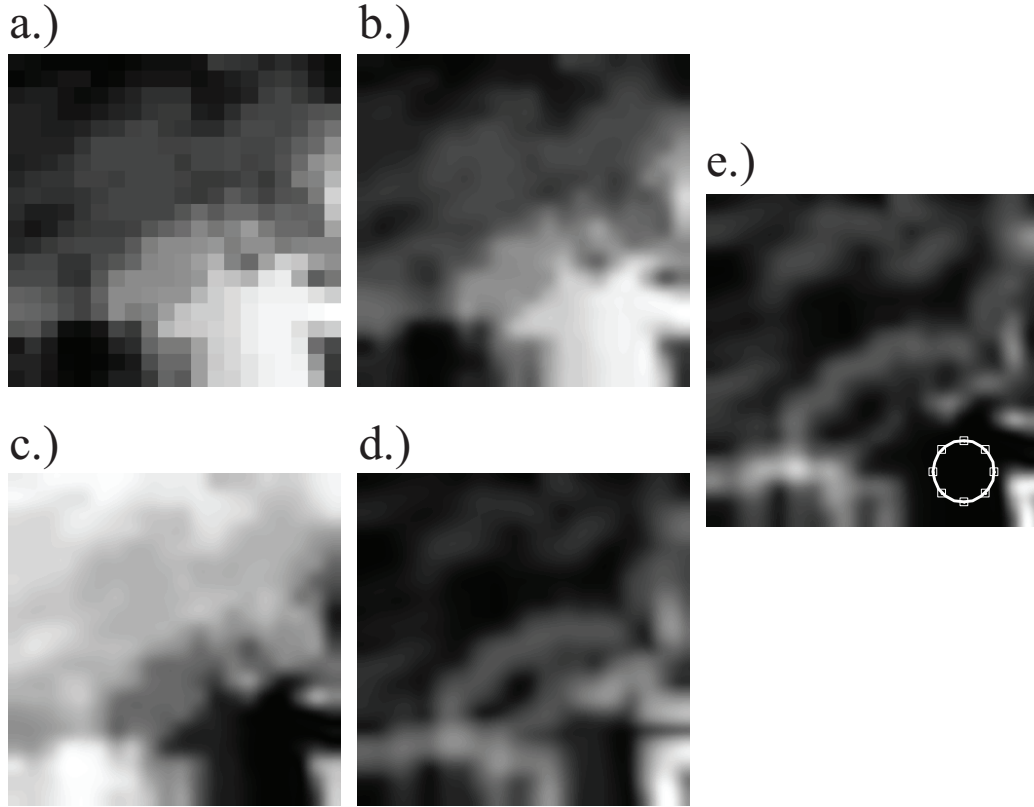


Figure 3.2: **Steps to determine the in-focus area for the first image in the stack.** a) The original 20x20 pixel depth map was (b) up-sampled to 1080 pixels and smoothed with a Gaussian, so that (c) the distance in depth of each pixel from the first image's focal depth could be computed. d) The local standard deviation of the depth image, computed over a 165x165 pixel region, was used to identify areas of uniform depth. e) The product of c and d: the white circle is centered on the location where this product has the lowest value (closest depth and lowest variance), to indicate the position of the in-focus patch at this depth.

value from the depth values in the up-sampled and smoothed depth map was computed as a  $1080 \times 1080$  matrix (Figure 3.2c). This matrix was then multiplied by the local standard deviation (computed over a  $165 \times 165$  pixel region) of the up-sampled depth map in order to give greater weight to the parts of the image which had more homogeneous depth distributions (Gonzalez and Woods 2008), as shown in Figure 3.2d. The minimum of the product of the depth and depth variance matrix was taken to be the pixel coordinates in which the image was in-focus (Figure 3.2e). Figure 3.3 shows an example of the results of this method. The left image shows the in-focus patch identified for the first image of the stack. The right image shows the locations on the depth map where in-focus patches were found for all 11 images in the stack.

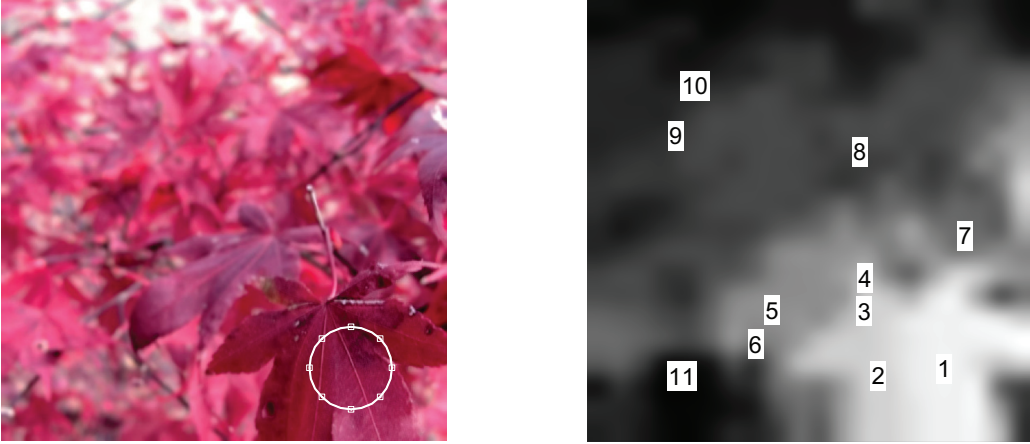


Figure 3.3: **Example of extracted in-focus patches for each focal plane.** Left: in-focus patch for the first image in the stack. Right: locations of in-focus patches on the depth map.

The Lytro camera captures light field pictures using a micro-lens array

between the image sensor and the main lens. The technical details are extensively explained in Ng (2006). In a conventional lens system, blur can be defined as the diameter of the circle  $C$  over which the point at distance  $Z$  is imaged on the image sensor placed at distance  $s$  from a lens with focal distance  $f$  and aperture  $A$  (Figure 3.4).  $C$  can be defined as (Held et al. 2010):

$$C = As \left| \frac{1}{f} - \frac{1}{Z} \right| \quad (3.1)$$

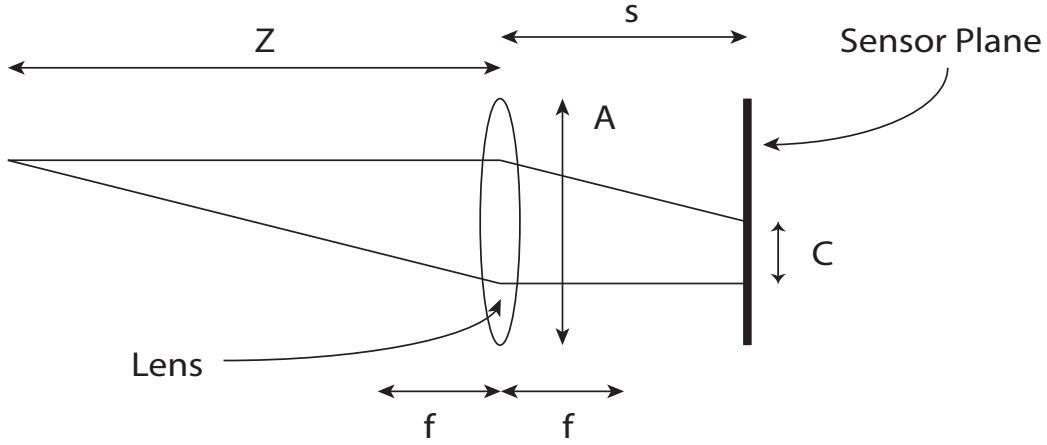


Figure 3.4: **Blur in a conventional lens system.** An object at distance  $Z$  from a lens with focal length  $f$  and aperture  $A$  is imaged on a sensor plane at distance  $s$ . The object will be blurred over a circular region with diameter  $C$ .

The Lytro software outputs depth information in lambda units. One lambda is the distance the sensor plane of a conventional camera must move to change the diameter of the circle of confusion  $C$  by the pitch of the micro-lens array. The pitch of the micro-lens array of the camera employed was

specified to be  $13.9 \mu\text{m}$ , the focal distance was 6.5 mm, and the aperture was  $f/2$ . The circle of confusion when moving the sensor plane from distance  $f$  (focus at infinity) to  $s$  is given by:

$$C = \frac{As}{Z} \quad (3.2)$$

By using the thin lens equation and rearranging, object displacement in diopters can be expressed as

$$d_0 = \frac{1}{f} - \frac{1}{s} \quad (3.3)$$

where the displaced focal plane  $s$  is given by:

$$s = \frac{f(A + C)}{A} \quad (3.4)$$

Through these calculations it is possible to recover the dioptric distance at which each image in the stack was in focus. Figure 3.5 shows the mean distance for the images in the stack across all ten light field photographs in the study. The mean dioptric distance between two image planes was 0.89 diopters.

In the Experiments, a Reference and a Test image patch were selected from differing depth planes in the same scene. Reference and Test stimuli were presented in circular windows, with a diameter of 6 degrees and whose edges were smoothed over 0.6 degrees with a raised cosine envelope. Subjects judged the relative depth of Reference and Test patches. There were four conditions in which stimuli contained differing sources of depth information:

1) Pictorial: Reference and Test patches were in-focus with zero disparity. Observers could only employ naturally-occurring pictorial cues to perform

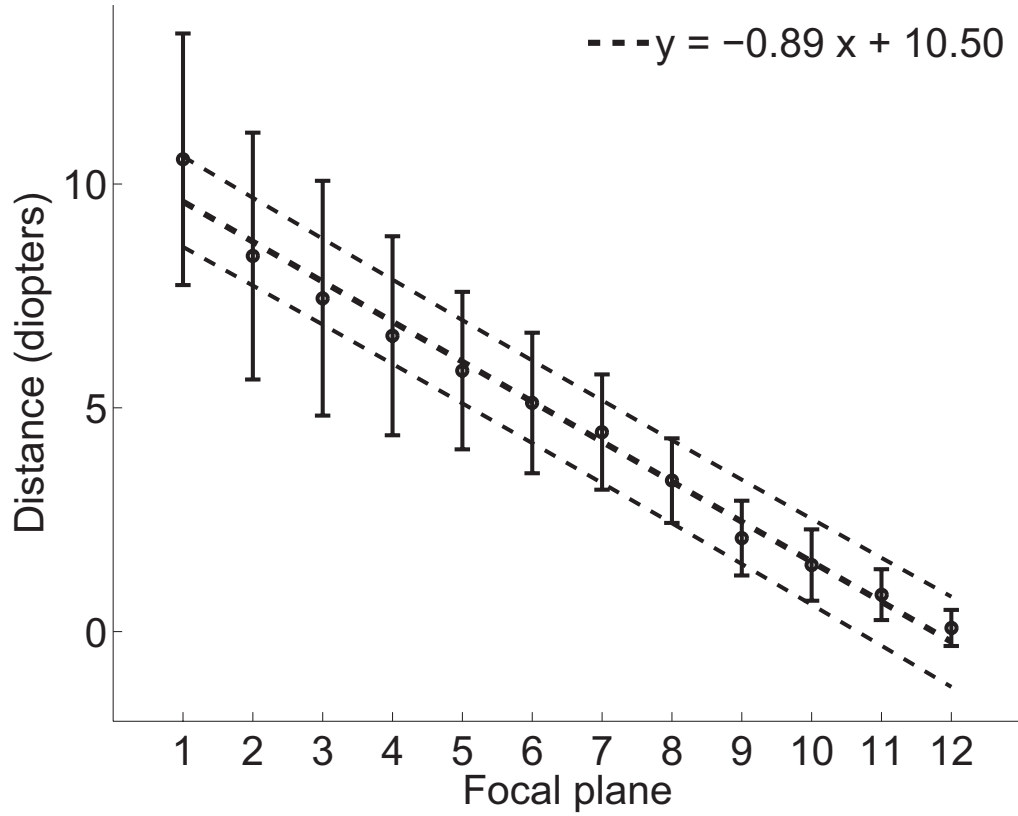


Figure 3.5: **Dioptric depth separation between image planes of light field pictures.** Results are for the ten light field pictures employed in the study. Black circles represent the mean dioptric distance for each image in the stack. Error bars are  $\pm 1$  standard deviation. Dotted lines represent the linear fit (coarse dots) with 95% confidence intervals (fine dots).

depth ordering, since no depth information was present from disparity or defocus blur.

2) Pictorial plus Blur: in addition to any naturally-occurring pictorial cues, the Reference patch was in focus, to simulate accommodation at the plane of the reference image, while the Target patch was blurred proportionally to the depth separation between the reference and target patches. Both patches were taken from the same image in the Lytro stack, so that the patch that was in-focus simulated accommodation at that plane, and the patch that was out of focus simulated appropriate defocus blur for the relative depth of that plane. Image blur arose directly from the optics of the Lytro camera, and was not added via image processing.

3) Pictorial plus Disparity: in addition to any naturally-occurring pictorial cues, Reference and Target patches were offset with appropriate disparity proportional to the depth separation between the patches. Both stimuli were in-focus. The Reference patch was presented at zero disparity, the Target disparity was crossed or uncrossed according to the distance in depth between the patches. The absolute values of the disparities ranged from 4 to 40 arcmin.

4) Pictorial plus Blur and Disparity: pictorial, blur and disparity depth cues were all present and coherent. The Reference patch was presented in focus at zero disparity, the blur and disparity of the Target patch were proportional to the separation in depth between the patches.

## **Procedure**

### **Temporal 2AFC Procedure**

The sequence of events from a typical trial is shown in Figure 3.6. Each trial started with the presentation of a central fixation cross on a grey background. After 1 second, two circular image patches were shown for 200 ms, separated by a 500 ms inter-stimulus interval in which the fixation cross was again displayed. Observers were required to indicate by pressing one of 2 buttons, which of the two image patches presented in the two intervals was further. Feedback was given by a green fixation cross flash for a correct answer or red for an incorrect answer.

There were 10 repetitions for each of the 10 depth pairs for each of the four conditions. Thus each session consisted of 400 trials, 100 trials per depth cue condition. Each subject completed each session twice, for a total of 800 trials per subject. The presentation order of the Reference and Test stimulus and the order of the presentation of the conditions was random.

### **Spatial 2AFC Procedure**

In the temporal 2AFC procedure, the visual targets containing blur and disparity appeared at fixation and might have induced vergence and possibly accommodation away from the fixation plane. Additionally, presenting blur and disparity at the center of the visual field is not typical for real world conditions where binocular fixation projects focused images with zero disparity on the fovea. Objects nearer or farther than the plane of fixation will instead be projected onto the peripheral visual field, at different retinal disparities,

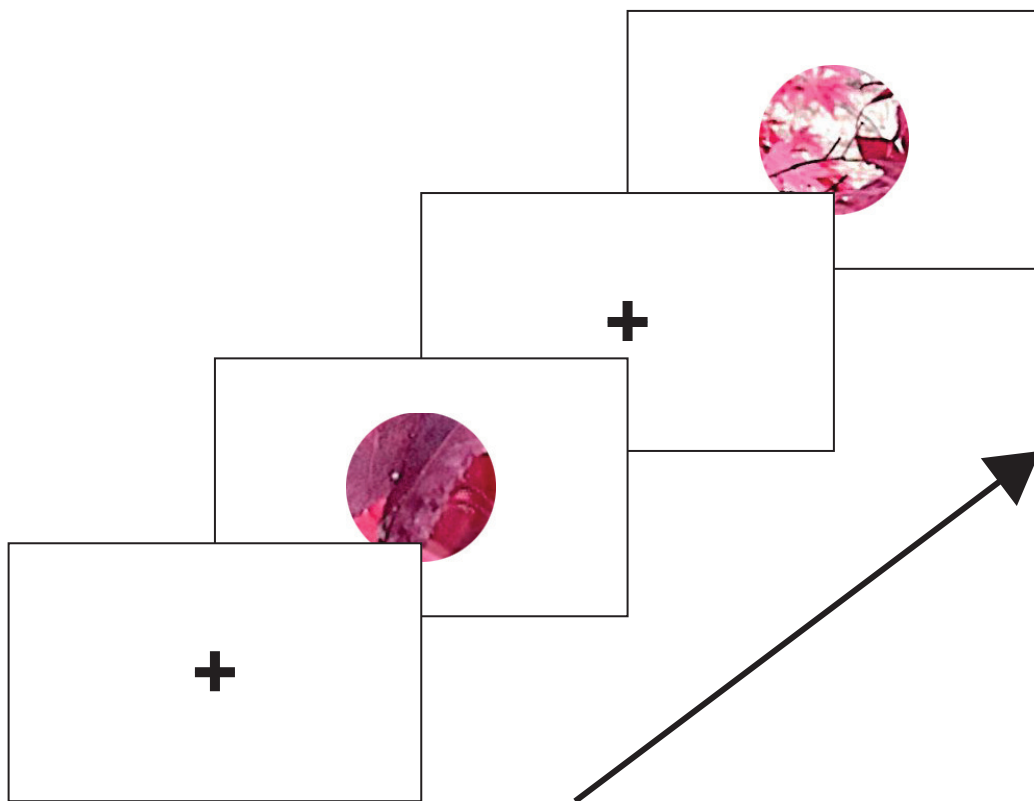


Figure 3.6: **Example of one trial from the temporal 2AFC experiment.** Each trial began with a blue fixation cross presented in the center of a grey screen for 1 s. Two stimuli were presented in a circular window with abrupt onset and offset for 200 ms, separated by a 500 ms interval containing only the fixation cross. The subject was required to fixate the cross and to indicate whether the first or second patch was farther, using any available cues.



and will be out of focus. I therefore repeated Experiment 1 except with a spatial paradigm to keep vergence and accommodation at the plane of fixation (the surface of the computer monitor), and simultaneously presented cues at other depth planes. The paradigm was the same as in Experiment 1, except that the reference and target patches were presented concurrently at 4 degrees to the left and right of fixation.

Each trial started with the presentation of a blue central fixation cross and white peripheral nonius lines on a grey background. The nonius lines served to aid vergence on the surface of the monitor. After 1 s, two circular image patches were shown for 200 ms (Figure 3.7). Subjects were instructed to fixate on the central cross and report, by pressing one of two computer keys, whether the left or the right image patch appeared farther. The fixation cross was green following a correct response or red following an incorrect response.

## Analysis

To measure discrimination sensitivity the data for each subject and condition were converted into discrimination  $d'$  (Wickens 2002). To compute group confidence intervals on  $d'$  measurements a bootstrapping procedure was employed (Efron and Tibshirani 1993). Mean  $d'$  for each observer and each condition were computed from the original data resampled with replacement 5000 times. These bootstrapped distributions were then collapsed across observers to obtain group distributions for each condition. The group distributions were fit with a Gaussian function from which the 2.5th and 97.5th quantiles were taken as 95% confidence intervals. Mean and standard deviation of fitted group distributions were employed to Z-transform

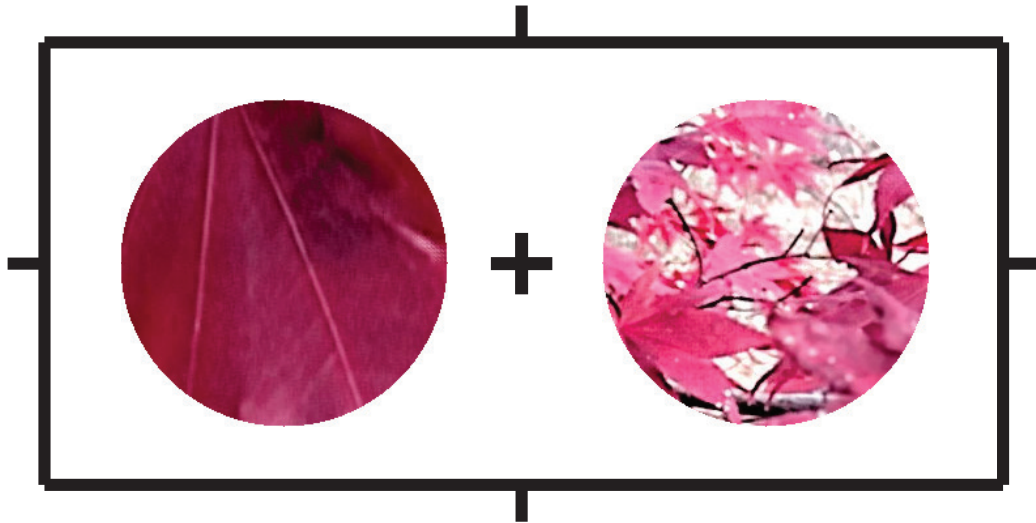


Figure 3.7: **Example of one trial from the spatial 2AFC experiment.** While fixating on the central cross, a subject was required to indicate which patch, the left or the right, was farther, using any available cues.

group means and compute p-values. Group averages for different conditions that fell outside the bounds of each other's confidence intervals (corresponding to p-values  $< 0.05$ ) were considered to be significantly different from one another.

### 3.3 Results

#### Experiment 1: Temporal 2AFC Procedure

Figure 3.8a shows discrimination  $d'$  for the four experimental conditions averaged across observers. When only pictorial cues to depth were present, discrimination was already possible ( $d' = 1.56$ , 95% confidence intervals [1.39

1.78]). When blur cues were introduced, discrimination significantly worsened ( $d' = 1.22$ , 95% confidence intervals [1.04 1.43],  $p=10^{-3}$ ). When disparity cues were introduced alongside pictorial cues, discrimination improved significantly ( $d' = 2.14$ , 95% confidence intervals [1.91 2.40],  $p=10^{-7}$ ). When both blur and disparity cues were present alongside pictorial cues, discrimination was again significantly better than with pictorial cues alone ( $d' = 1.94$ , 95% confidence intervals [1.75 2.23],  $p=10^{-3}$ ), and not significantly different from the pictorial + disparity condition ( $p=0.11$ ).

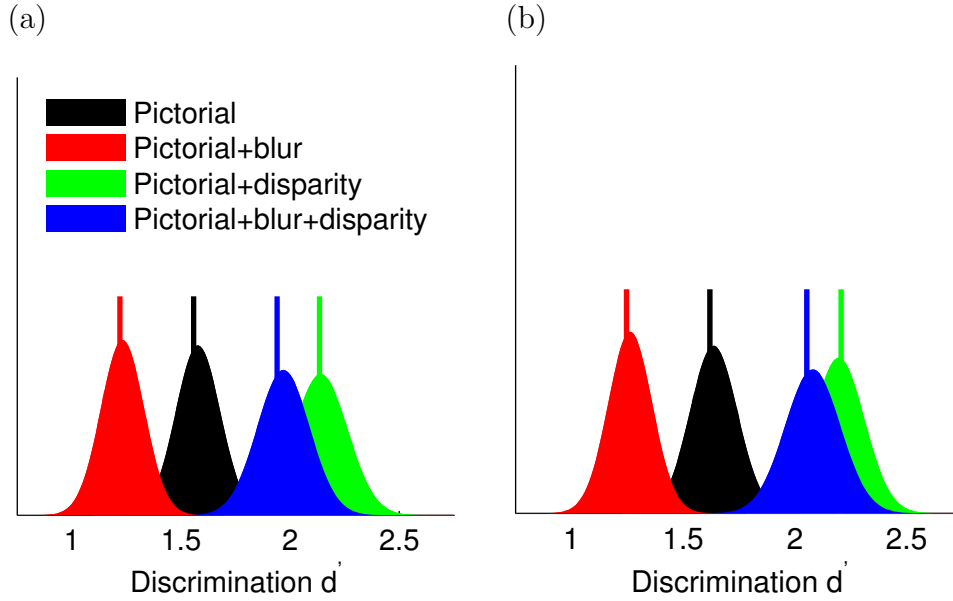


Figure 3.8: **Depth discrimination sensitivity.** Discrimination sensitivity  $d'$  for each condition in the (a) temporal 2AFC task and (b) spatial 2AFC task, averaged over the four observers in each experiment. The graphs show group bootstrapped distributions of discrimination  $d'$  for the four experimental conditions. Vertical bars represent the true group means.

As expected, depth order discrimination was already possible solely with pictorial cues to depth. Binocular disparity, which is a robust cue to depth perception, significantly improved discrimination sensitivity. Contrary to expectation, the addition of blur cues to depth impaired depth order discrimination, probably by reducing the contrast of pictorial depth cues at high spatial frequencies. However, blur did not significantly impact depth order discrimination due to binocular disparity, suggesting that disparity is a robust cue to depth even with low spatial frequency information.

## Experiment 2: Spatial 2AFC Procedure

Figure 3.8b shows discrimination  $d'$  for the four experimental conditions averaged across observers. When only pictorial cues to depth were present, discrimination was possible ( $d' = 1.62$ , 95% confidence intervals [1.43 1.85]), confirming the results of Experiment 1. When blur cues were introduced, discrimination significantly worsened ( $d' = 1.25$ , 95% confidence intervals [1.07 1.46],  $p=10^{-4}$ ). When disparity cues were introduced alongside pictorial cues, discrimination improved significantly ( $d' = 2.21$ , 95% confidence intervals [1.96 2.43],  $p=10^{-7}$ ). When both blur and disparity cues were present alongside pictorial cues, discrimination was again significantly better than with pictorial cues alone ( $d' = 2.05$ , 95% confidence intervals [1.84 2.34],  $p=10^{-4}$ ), and not significantly different from the pictorial + disparity condition ( $p=0.19$ ).

For both experiments, I repeated the analysis using conventional detection thresholds, fit with logistic psychometric functions, and found a similar pattern of results ( $r^2 = 0.86$ ,  $p = 10^{-6}$  for Experiment 1;  $r^2 = 0.90$ ,  $p = 10^{-7}$

for Experiment 2). I however choose to present the  $d'$  analyses because it provides a more appropriate comparison of discrimination performance between experimental conditions.

### 3.4 Discussion

For binocular organisms with well-developed visual cognition, multiple cues are available in natural environments to indicate the absolute or relative depth of different objects. While combinations of these alternate cues have been studied with simplified laboratory stimuli, there have been relatively few attempts to examine depth cue combination in natural scenes. I used a plenoptic camera to capture light field images of real scenes, which allowed us to control and combine depth cues from pictorial, binocular disparity and defocus blur sources. I found that even when pictorial information was impoverished by a small field of view, partial occlusion and parafoveal viewing, pictorial cues alone enabled relative depth judgments. Stereoscopic disparity provided significant benefit for depth perception both in the near periphery and at the fovea. Contrary to recent data (Held, Cooper, and Banks 2012), I find that the addition of defocus blur cues to depth impaired depth perception.

Landy et al. (1995) propose a model of depth cue combination termed modified weak fusion (MWF), which consists of a dynamically weighed average of available cues, based upon the estimated reliability of the cues. In this study I found that depth order discrimination was already possible with pictorial cues alone. This supports previous evidence that pictorial cues such as

texture perspective are strong cues to depth that can be used in combination with other depth cues and can at times take precedence over other depth cues such as motion parallax (Young, Landy, and Maloney 1993; O’Brien and Johnston 2000).

Held, Cooper, and Banks (2012) have recently shown that depth discrimination thresholds for random dot patterns decrease when blur is varied in proportion to depth. I found instead that adding dioptric defocus blur cues to depth actually impaired depth perception. The main difference between this study and the previous is that Held et al. employed a volumetric display to present observers with blur due to the optics of their own eyes. Thus, in the Held study, observers might have been able to employ other optical cues, such as chromatic aberration, that have been shown to be possible cues to depth sign (Nguyen, Howard, and Allison 2005). Furthermore, in the study by Held, Cooper, and Banks, an increase in blur was always correlated with an increase in depth away from fixation, thus the observers could perform the task using blur discrimination (the decrease in depth was pilot tested but the data were not presented). In the present study, blur increased with distance away from the plane of focus, regardless of the sign of the distance (i.e. closer or further).

The addition of blur cues impaired depth discrimination by reducing the contrast of geometric and disparity information at high spatial frequencies. According to previous work, subjects should have been able to perform depth ordering using image defocus blur as a cue to depth (Nguyen, Howard, and Allison 2005; Held, Cooper, and Banks 2012) by solving the sign problem using pictorial cues. Although the amplitude of contrast attenuation from

defocus blur is an unsigned quantity, relative phase differences between near and far objects of equal blur could, in principle, be used to estimate the sign of depth difference. My results instead suggest that subjects were unable to use generic defocus blur in the optics of the Lytro camera as a cue to depth.

The results from the spatial 2AFC experiment confirmed the results from the temporal 2AFC paradigm. Accommodative response to blur did not seem to influence the results, possibly because accommodative latency is greater (300 ms) (H. A. Anderson et al. 2010) than the stimulus duration employed here (200 ms). Fusional vergence during the 2AFC temporal experiment could have affected the stimulus, since the typical vergence onset latency (approximately 160 - 180 ms) (Schor and Ciuffreda 1983) is less than the stimulus duration employed. In the temporal 2AFC paradigm this could have lead to improved performance in the stereoscopic conditions by added depth information from oculomotor vergence cues with respect to the spatial experiment. The fact that there were no observable differences between the paradigms supports previous evidence that stereopsis is a direct process (Lugtigheid et al. 2014), where depth is estimated from retinal disparity and not from fusional vergence responses.

Since the defocus blur depth cue did not arise from the optics of an individual's eye, it is possible that subjects could learn the correct association between defocus blur of the Lytro camera and object depth, in the same way that they may have learned this relationship for their own optics. In this study, feedback was given to verify whether subjects could learn that association, yet there was no evidence of perceptual learning in either the temporal nor the spatial experiment.

To estimate depth discrimination sensitivity, data were pooled over all depth differences. Because sensitivity to different cues may change with depth difference, it remained possible that blur might have had an adverse effect at some depth differences, and a positive effect on depth discrimination at others. However, there was no evidence of an association between sensitivity to the different cues and the magnitude of the depth differences.

Furthermore, I considered the possibility that blur may have had different effects at far or near absolute distances. To address this question, I fit raw discrimination data with asymmetric logistic functions and compared the symmetry of the fitted functions. I found no evidence that blur asymmetrically modified depth discrimination due to perspective (symmetry of pictorial vs symmetry of pictorial+blur conditions:  $p=0.12$  for Experiment 1;  $p=0.26$  for Experiment 2, paired samples t-test) or disparity cues (symmetry of pictorial+disparity vs symmetry of pictorial+disparity+blur conditions:  $p=0.31$  for Experiment 1;  $p=0.19$  for Experiment 2, paired samples t-test). This suggests that pooling the data did not mask conditions in which blur is beneficial. Instead, the results indicate that in naturalistic images with signed depth, image blur does not facilitate depth discrimination.

These observations confirm the conclusion that generic defocus blur is unlikely to be successful at facilitating depth perception in naturalistic images. This result has important implications for virtual reality applications. I propose that for virtual reality applications, the addition of dioptric blur, which can provide usable cues to depth in laboratory settings (Marshall et al. 1996; Mather 1996; Mather and Smith 2000; Nguyen, Howard, and Allison 2005; Hoffman and Banks 2010; Held, Cooper, and Banks 2012; Vish-



wanath 2012), is too unreliable when presented with other, more reliable, depth cues. Consequently, generic defocus blur may be negligibly weighted by the visual system in the recovery of depth information in naturalistic images that contain multiple potential depth cues. This conjecture is supported by the observation that depth discrimination was best when geometric and stereoscopic disparity cues were both present. This indicates that these cues can be combined, as previously reported by Johnston, Cumming, and Parker (1993) and Landy et al. (1995) who have proposed weighted cue combination rules. When Johnston, Cumming, and Parker (1993) introduced blur cues along with geometric and disparity cues, they marginally impaired discrimination, in agreement with my observations and suggesting that disparity and pictorial cues are weighed stronger than defocus blur. Furthermore this result is in line with the notion that disparity can be used even with low spatial frequency information (Schor, Wood, and Ogawa 1984a) that is encoded by coarse disparity detectors (Felton, Richards, and Smith 1972; Marr and Poggio 1979) and still produce stereoscopic depth percepts.

The main implication of my findings with regards to virtual reality technology is that rendered generic defocus blur will not necessarily facilitate the fine perception of relative depth in complex 3D stimuli. Furthermore, applying blur to pictorial depth cues may even reduce their effectiveness, especially when information is carried by high spatial frequencies, and this detrimental effect might hinder depth cue combination. Nevertheless, blur might play other roles in virtual reality rendering techniques, as other studies have found that rendered blur has an array of effects on the perceptual experience of virtual reality.

Vishwanath and Blaser (2010) have shown that blur gradients, which can be presented as independent from the depth structure of the rendered scene, can modulate the perception egocentric distance, and play a role in the tilt-shift miniaturization effect. Similarly, Wang et al. (2011) provided evidence that a globally blurred background enhances the perceived depth separation between the background and a sharp, disparity defined foreground object. Nefs (2012) found that depth of field systematically affected perceived depth/width ratio of photographs of natural scenes, while Zhang et al. (2014) have shown complex patterns of dependencies between depth of field, height-in-the-field, 3D display system, and perceived depth when observers were asked to draw floor plans of viewed scenes to scale. All these studies find that blur modulates the perception of global scene depth. Thus blur might still be gainfully employed in virtual reality applications to modulate egocentric distance and global scene appearance (much like what is done in cinematography (Katz 1991; Kosara, Miksch, and Hauser 2001)), and to potentially facilitate oculomotor behavior. In the next chapter I specifically test whether rendering blur can produce benefits on binocular fusion of stereoscopically rendered naturalistic scenes.

## Chapter 4

# Simulated Disparity and Peripheral Blur Interact During Binocular Fusion

### 4.1 Introduction

The visual experience when using man-made entertainment devices (computers, television, books, newspapers) is very different from natural viewing conditions. While natural conditions contain a broad and continuously varying range of visual depth information, man-made displays usually contain a narrow range of fixed depth. For example, many artificial images contain little or no blur because they are designed for a user to be able to extract information from all areas of the image (e.g. when using a desktop application or when playing a videogame). When reading a book or using an e-reader the accommodative distance is much closer to us than in typical natural

conditions, especially when outdoors (Geisler 2008). In directed movies and television, spatial blur is often manipulated to induce the viewer to attend to specific, in-focus portions of the scene (Katz 1991; Kosara, Miksch, and Hauser 2001). Such conditions differ from the real-world - where the ranges of depths at which our eyes accommodate produce variation in retinal blur across space and over time. Furthermore, stereoscopic 3D is increasingly used to promote the illusion of depth, but current display technology nearly always presents depth information uncoupled from focus information.

The presence of defocus blur has been shown to diminish visual fatigue while viewing stereoscopic 3D stimuli (D. M. Hoffman et al. 2008). I was interested in simulating, via image processing, the changes in peripheral blur due to naturalistic accommodative changes. My aim was to develop a novel method of presenting both depth and blur information in a way that simulated natural viewing conditions. I implemented a real-time gaze-contingent stereoscopic display with naturalistic distributions of blur and disparity across the retina. At fixation, the display is kept in focus, analogously to foveation and accommodation in the real world. In peripheral vision, images are presented with varying amounts of blur, increasing with distance from the simulated depth plane of fixation. To control dioptric blur, as in the previous chapter I took advantage of Light Field Rendering photographic technology (Ng et al. 2005; Ng 2006). I employed light field photographs of natural scenes taken with the Lytro plenoptic camera (Figure 3.1). Light field cameras output for a single photographic exposure an image stack with images focused at different depths, a depth map of the captured scene (see Figures 3.1 and 3.3), and a depth look-up table. Although

this system does not approximate the optical aberrations of an individual, it provides a general purpose method to simulate changes in blur and depth that may be implemented in a practical, low cost system. In this chapter, I examine whether any functional benefit can be obtained with such a general approach.

While subjects freely viewed such photographs, I employed an eye tracker to monitor the point of fixation and to determine the depth of this location from the light field image's depth map. With this implementation, I then rendered the appropriate image dioptric blur for all other points in the image in real-time by selecting the appropriate image from the light field image stack. Thus, I allowed for the distribution of blur across the retina to be controlled in real time, which simulated a spatial and temporal distribution of image blur similar to that which occurs under free viewing in natural conditions. I used this display to examine how eye movements and binocular fusion depend on depth cues from blur and stereoscopic disparity in naturalistic images.

## **4.2 Methods**

### **Subjects**

Six subjects (aged from 21 to 33 years) completed the experiment, one of whom (GM) was the author of this work. All subjects had normal or corrected to normal acuity and normal stereo vision.

## Apparatus

Software was written with Matlab version R2011a using Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) interfaced via the EyeLink Toolbox (Cornelissen, Peters, and Palmer 2002) with an EyeLink 1000 Desktop Mount eye tracker with a 2000 Hz camera upgrade. Stimuli were presented on a Samsung SynchMaster 2233 LCD monitor at 120 Hz, run from an NVidia Quadro FX580 graphics processing unit that was positioned 50 cm from the observer. Display dot pitch was 0.282 mm. The spatial resolution of the stimulus images was 1080 by 1080 pixels and subtended 34 degrees of visual angle on the screen. Stereoscopic disparity was presented via the NVIDIA 3D Vision kit. The eye tracker was calibrated with a supplied 5 point calibration routine. The system latency was measured at  $17.5 \pm 2$  ms using the methods described by Saunders and Woods (2014). The measured latency falls well within the time frame in which visual sensitivity remains reduced following a saccade (Volkman 1962; Volkman, Schick, and Riggs 1968; Volkman et al. 1978) and thus system latency is unlikely to affect the presented results.

## Stimuli

Light field photographs of natural scenes were taken with a Lytro camera. I collected 30 photographs of natural scenes, selected to contain a distribution of objects across a range of depths. This selection process was arbitrary, but deliberately avoided panoramic or empty scenes that contained few or distant objects. I positioned the camera at eye level at locations where the

closest object was approximately 10 cm from the lens.

As mentioned in Chapter 3, the Lytro is a low cost plenoptic camera that allows the user to refocus on different parts of a picture after the picture has been captured. It was chosen for this work because it allows the extraction of multiple versions of the same image, each focused at a different depth plane. This allowed me to assemble a database of pictures of natural scenes where different objects could either be in focus or at an increasing level of blur with their distance in depth from the current plane of focus. I used open source software (Patel 2012) to extract data and images from the Lytro file format. This tool retrieved the component .jpg images at different focal depths, a depth map of the photographed scene, and a depth look up table, but did not apply any image processing of the files. The depth map is output with a resolution of 20 by 20 pixels, which I bicubically interpolated to 1080 by 1080 pixels to match the resolution of the .jpg images. I employed this depth map, which was output from the Lytro proprietary software, for both the gaze-contingent implementation and for stereoscopic rendering. To validate the use of the Lytro depth maps for these purposes, I correlated the depth maps of the 30 photographs employed with the corresponding sharpness maps obtained through the focus stacking technique described below, and found a mean correlation coefficient of  $r = 0.96$ , 95% CI [0.92, 0.98] (mean correlation coefficient and confidence intervals were estimated using Fisher's Z Transform). This suggests good agreement between the depth maps output by the Lytro proprietary software and the geometric and depth structure of the photographed images.

The image stack extracted from the Lytro data files contained from 1 to

12 images, depending on the depth structure of the scene. If a scene contains many objects at near depths to the camera, there will be more output images. If instead a scene is panoramic, and everything is focused at infinity, only one image will be generated.

## **Gaze Contingent Blur Implementation**

Each video frame, the gaze position in screen pixel coordinates was determined from the eye tracker. The depth of the pixel at the gaze coordinate was indexed from the depth map image. The absolute difference between the value of the depth map at the gaze coordinate and the focal depth of each of the plenoptic image planes was computed. The image for which this difference was minimum was selected as the image that was most in-focus at the current gaze position and was displayed to the screen. The viewer thus always viewed an in-focus image at the center of his/her visual field. Objects not in the center of the viewer's visual field but at the same focal depth as the fixation coordinates were also in-focus, while everything else in the image was optically blurred, with blur increasing with distance in depth from the plane of fixation.

## **Focus stacking**

To evaluate the effects of gaze-contingent blur, I compared the gaze-contingent display with a display where everything was presented in sharp focus. The stack of images extracted from a photograph taken with the Lytro was used for this purpose through focus stacking. Each of the images in the stack is in-focus in different regions of the scene. To create an image that is fo-



cused everywhere, the in-focus patches in each of the images at different focal depths were located and the in focus patches were blended together to generate a final image.

In order to identify the most focused image for each region, I assumed that the focused regions of an image would contain more sharp edges than the same region in other images (Gonzalez and Woods 2008). A 3x3 pixel Laplacian filter was used to locate sharp edges for each image. The edge map computed for each image was then smoothed with a 35x35 pixel moving average filter to indicate the sharpness level of each pixel. Each pixel in the final stacked image was selected from the image with the highest average value at those pixel coordinates. Figure 4.1 shows an example of the results of focus stacking, which produced an image that is focused all over.

## **Stereoscopic Images from Light Field Data**

The images extracted from the plenoptic camera are not intrinsically stereoscopic, but I used the depth map to create stereoscopic image pairs for each photograph. I created a stereoscopic image for the left and right eye by shifting individual regions of the original image by distances in proportion to the depth value given by the depth map (Figure 4.2). Each pixel pair from the original image was shifted  $d/2$  pixels to the left and  $d/2$  pixels to the right in the stereo image pair, where  $d$  defines the stereoscopic disparity for the depth value of the original pixel. Occluded pixels in the farther depth plane were completed with the value of the closest unoccluded pixel to generate an abrupt depth step between foreground and background regions. These interpolated pixels were not easily detected owing to strong crowding effects in

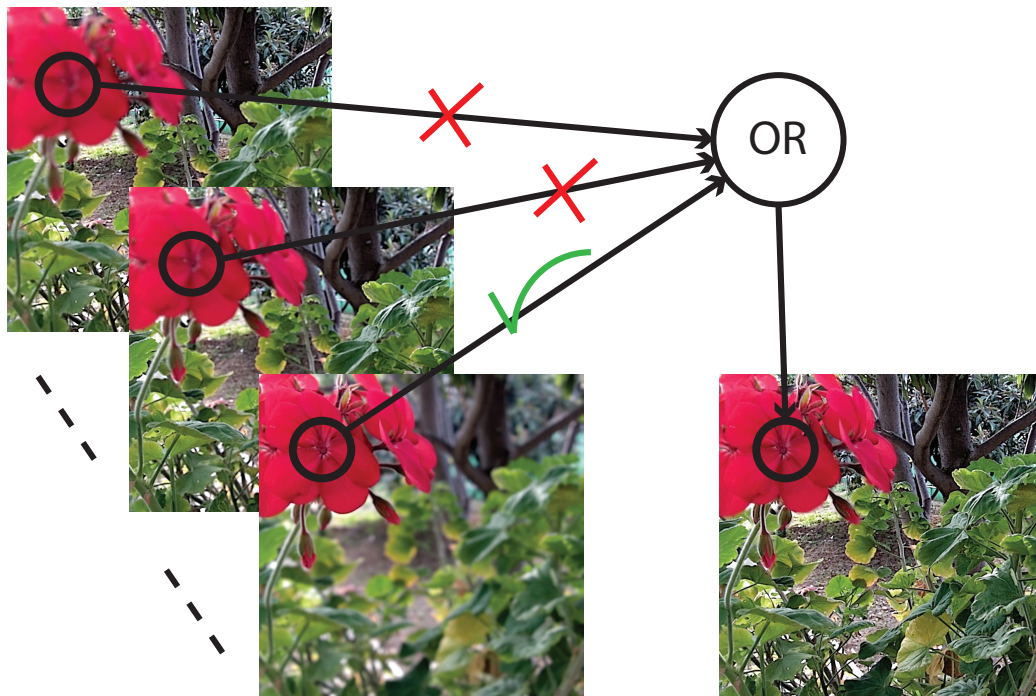


Figure 4.1: **Focus stacking.** In each image extracted from the Lytro camera different areas will be in focus. The most in-focus portions of each image will be automatically selected to slice together a final image which is focused all over.

natural scenes (Balas, Nakano, and Rosenholtz 2009; Wallis and Bex 2012).

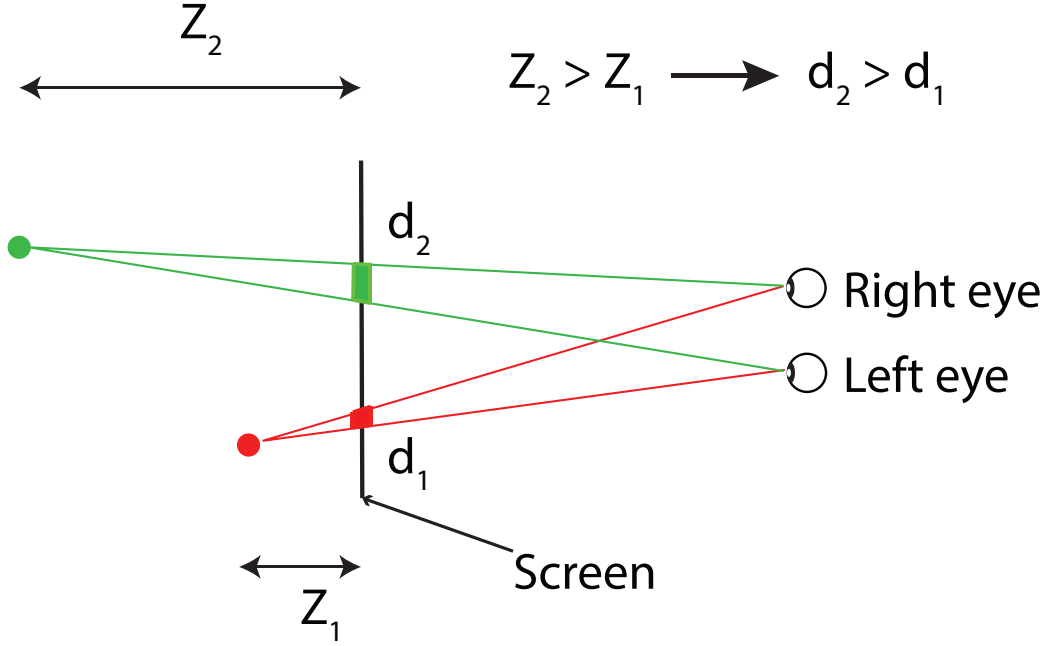


Figure 4.2: **Geometry of disparity rendering.** A point at distance  $Z_1$  is rendered on the monitor with disparity  $d_1$ . A point at a greater distance  $Z_2$  from the monitor is rendered on the surface of the monitor with greater disparity  $d_2$ . Objects more distant from the surface of the monitor are rendered on the surface of the monitor with greater disparities with respect to objects that are closer in depth to the surface of the monitor.

Next, the the maximum crossed and uncrossed disparity values were specified; for example specifying the disparity range from 0 to 125 pixels, and given the dot pitch size of 0.282 mm, we obtain a retinal disparity ranging from 0 to 4 degrees at the 50 cm viewing distance. In this configuration the closest objects would be rendered on the surface of the screen and the

farthest objects would be rendered approximately 60 cm behind the surface of the screen. Images containing crossed disparity had maximum values of disparity of either -2.4 degrees or -5.7 degrees. In these images the farthest object in the scene had zero disparity and was perceived on the surface of the screen, while the nearest objects should appear in front of the screen. Images containing uncrossed disparity had maximum values of disparity of either 2.4 degrees or 5.7 degrees. In these images the closest object in the scene had zero disparity and was perceived on the surface of the screen, while the farthest objects should appear behind the surface of the screen. Stereoscopic image pairs with all four possible ranges were generated for each of the 30 images employed so that disparity was not confounded with image identity. The stereoscopic image pair was presented through shutter glasses. These large stereoscopic disparities were employed to test the limits of stereoscopic fusion while observers were allowed to view stereoscopic stimuli for prolonged periods of time and were free to make multiple saccadic and vergence eye movements (Yeh and Silverstein 1990).

## Procedure

To evaluate this system, I examined how eye movements and binocular fusion depend on depth cues from blur and stereoscopic disparity in natural images. I measured the time to perceptual fusion for images in which blur was gaze-contingent compared to focused stacked images in which all points were in focus. Subjects were seated with their heads fixed in a chin and forehead rest at a distance of 50 cm from the monitor. All subjects wore NVIDIA 3D shutter glasses. Observers were asked to freely-view the images and to

report when the image appeared to fuse into 3D by pressing a button on a keyboard placed in front of them.

Figure 4.3 illustrates an experimental trial. Each trial began with subjects looking at a uniformly grey screen. A red fixation cross was presented at a location where the binocular disparity was greatest in the stimulus image to be presented. This ensured that before fusion was experienced, the observer should experience maximal diplopia at the fovea.

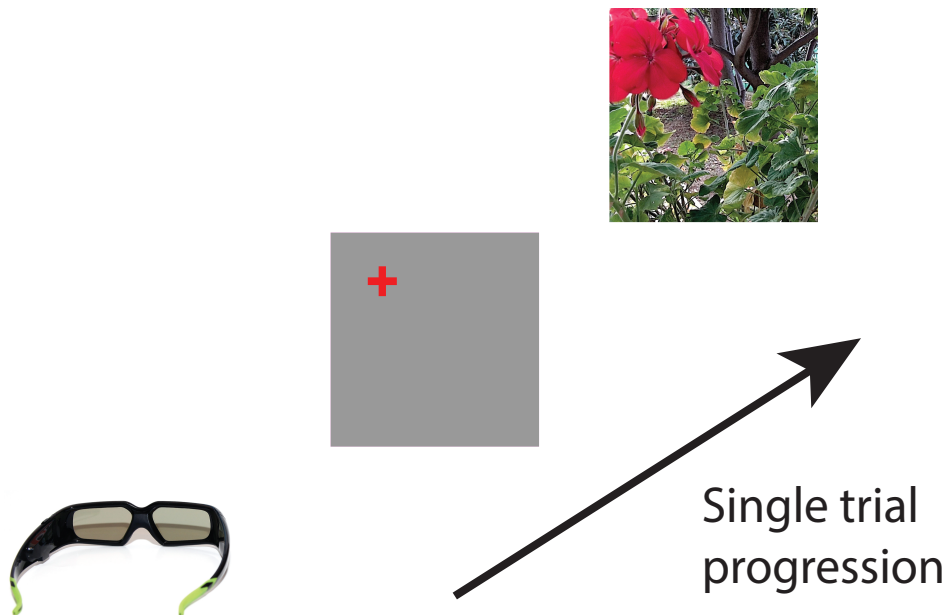


Figure 4.3: **Example of a binocular fusion experimental trial.** When subjects fixated on the red cross, the grey screen would be replaced by the stimulus image.

As soon as the eye-tracker data confirmed that the subject was fixating within 1.6 degrees of the center of the cross, the stimulus image was presented.

Subjects were instructed to report when the portion of the image initially cued was perceived as fused by pressing the space bar on the keyboard in front of them. A complete session consisted of 240 trials. Each of the 4 levels of disparity was presented 60 times in one session: in random order, 30 stimulus images contained gaze-contingent blur, 30 stimulus images were presented focused all over.

### 4.3 Results

All data was analyzed using a 2 (blur) x 4 (disparities) repeated-measures ANOVA. There was large inter-subject variability in the time to perceptual fusion. Figure 4.4a shows boxplots for the 6 observers, which show the mean (.) median (\*) interquartile range (box) 95% tails (whiskers) and outliers (o). The raw data were therefore normalized for each subject in order to perform comparisons of the effects of blur and disparity across observers. For each observer, the mean time to perceptual fusion in each condition was divided by the mean time across all conditions. The relative time to fusion across conditions was then averaged across observers.

Figure 4.4b shows the relative time to perceptual fusion as a function of the disparity presented (x axis) for gaze contingent blur (blue) and focused (red) images. ANOVA results showed a significant main effect of disparity,  $F(3,15)=24.68$ ,  $p<0.00001$ , no significant main effect of blur,  $F(1,5)=2.76$ ,  $p=0.16$ , and a significant interaction between blur and disparity  $F(3,15)=3.78$ ,  $p<0.05$ . Figure 4.4b shows a reduction in relative time to fusion at high crossed and uncrossed disparities for the gaze contingent blur condition with

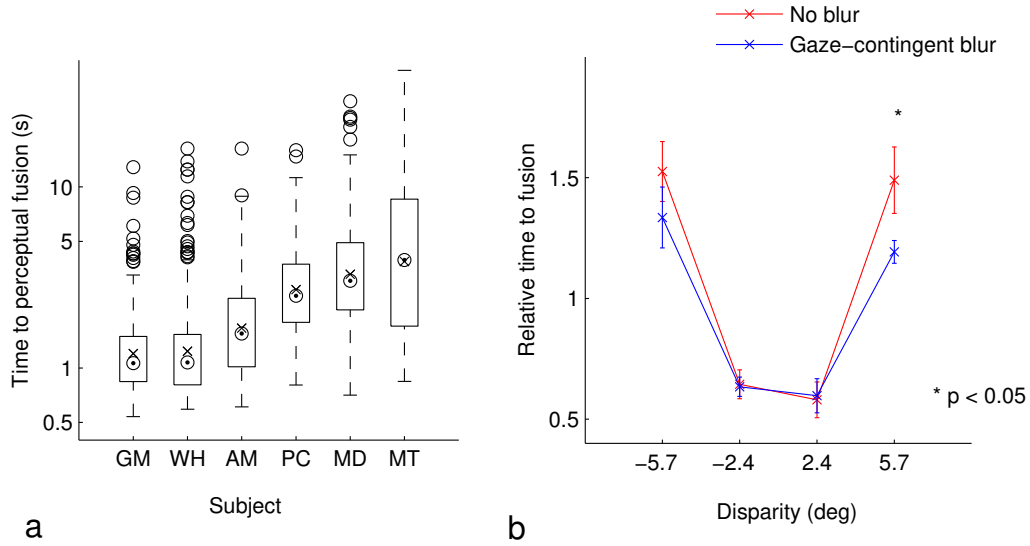


Figure 4.4: **Time to perceptual fusion.** a) Boxplot distributions of the time to perceptual fusion throughout the experiment for the 6 subjects. Dots are medians, crosses are means, box shows interquartile range, whiskers show 95% tails, and circles represent outliers. Note that the Y axis is log scaled. b) Relative time to perceptual fusion, normalized to the mean, as a function of disparity. The data are the mean for the 6 observers, error bars represent  $\pm 1$  SEM (standard error of the mean). The time to fusion in each condition was normalized to the mean for each observer across all conditions. The blue curve shows the gaze contingent blur condition, the red curve shows the focused everywhere condition.

respect to the focused images. Post hoc analysis showed a significant reduction in the time necessary to fuse stereoscopic images when informative distributions of retinal blur were present at high uncrossed disparities ( $p < 0.05$ , Bonferroni corrected t-test). The size of the reduction in time to fusion at high uncrossed disparities is highly correlated ( $r = 0.97$ ,  $p < 0.001$ ) with the time to fuse high uncrossed disparities with no gaze contingent blur. This suggests that the greatest benefits are gained by the subjects who take the longest to perceptually fuse stereoscopic images.

## Analysis of Eye Movements

Time to perceptual fusion and the number of eye movements in each condition were highly correlated ( $r = 0.92$ ,  $p < 10^{-20}$ ). As per the time to fusion data, there was large inter-subject variability in the number of eye movements made during each trial. Figure 4.5a shows boxplots for the 6 observers, which show the mean (.) median (\*) interquartile range (box) 95% tails (whiskers) and outliers (o). The raw data were therefore normalized for each subject, as per the time to fusion data. The relative number of fixations across conditions was then averaged across observers.

Figure 4.5b shows the relative number of fixations that were made before the images appeared to fuse as a function of the disparity presented (x axis) for gaze contingent blur (blue) and focused (red) images. ANOVA results showed a significant main effect of disparity,  $F(3,15) = 15.50$ ,  $p < 0.0001$ , no significant main effect of blur,  $F(1,5) = 6.36$ ,  $p = 0.05$ , and a significant interaction between blur and disparity  $F(3,15) = 5.90$ ,  $p < 0.01$ . Post hoc analysis showed a significant reduction in the number of fixations necessary to



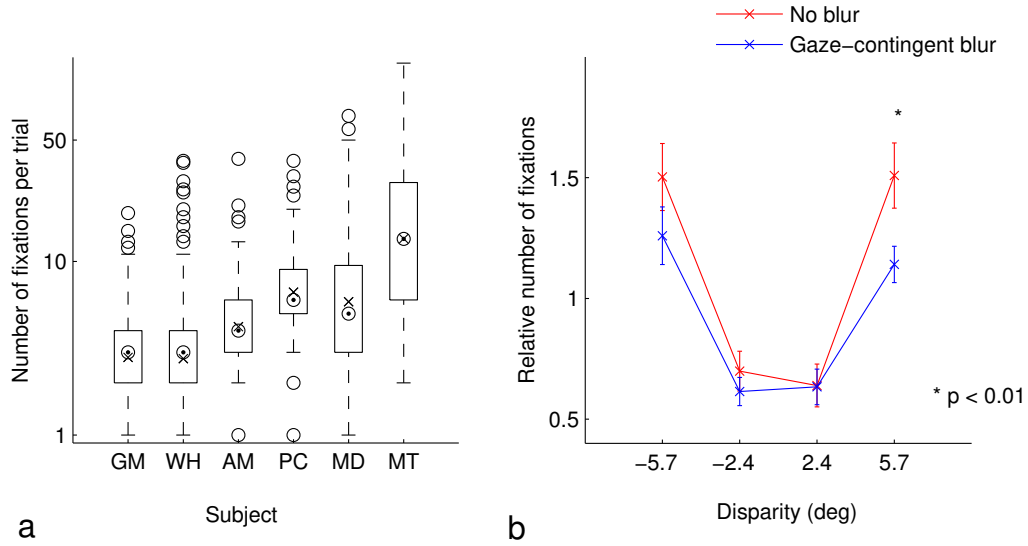


Figure 4.5: **Fixations necessary for perceptual fusion.** a) Boxplot distributions of the number of fixations per trial throughout the experiment for the 6 subjects. Dots are medians, crosses are means, box shows interquartile range, whiskers show 95% tails, and circles represent outliers. Note that the Y axis is log scaled. b) Relative number of fixations, normalized to the mean, necessary to fuse the images as a function of disparity, the data are the mean for the 6 observers, error bars represent  $\pm 1$  SEM. The number of fixations in each condition was normalized to the mean for each observer across all conditions. The blue curve shows the gaze contingent blur condition, the red curve shows the focused everywhere condition.

fuse stereoscopic images when informative distributions of retinal blur were present at high uncrossed disparities ( $p < 0.01$ , Bonferroni corrected t-test).

Fixation duration (Figure 4.6) was longer for gaze-contingently blurred images  $F(1,5)=28.40$ ,  $p < 0.01$ , but there was no significant effect of disparity  $F(3,15)=0.49$ ,  $p=0.69$  and no significant interaction between blur and disparity  $F(3,15)=0.24$ ,  $p=0.86$ .

When attempting to fuse a stereoscopic image, the eyes need to verge to the correct depth plane. I hypothesized that vergence stability would be elevated along the horizontal direction until stereoscopic fusion is achieved. To test this hypothesis, I computed horizontal vergence stability for each fixation, and compared it to vertical vergence stability. Horizontal and vertical vergence stability (SH and SV) were defined as the inverse of the standard deviation of horizontal and vertical vergence, which in turn were defined as the difference between the coordinates of each eye along the horizontal and vertical directions:

$$S_H = \frac{1}{std(x_L - x_R)}, S_V = \frac{1}{std(y_L - y_R)} \quad (4.1)$$

Figure 4.7 shows median horizontal (red) and vertical (black) vergence stability, averaged across the six subjects, for the first ten fixations in every trial.

The first fixation in each trial usually shows the lowest horizontal vergence stability (highest standard deviation in horizontal vergence). Along the horizontal direction, stability increases over the first few fixations, and plateaus in latter fixations. Vertical vergence stability was substantially greater than horizontal vergence stability, and remained approximately constant through-

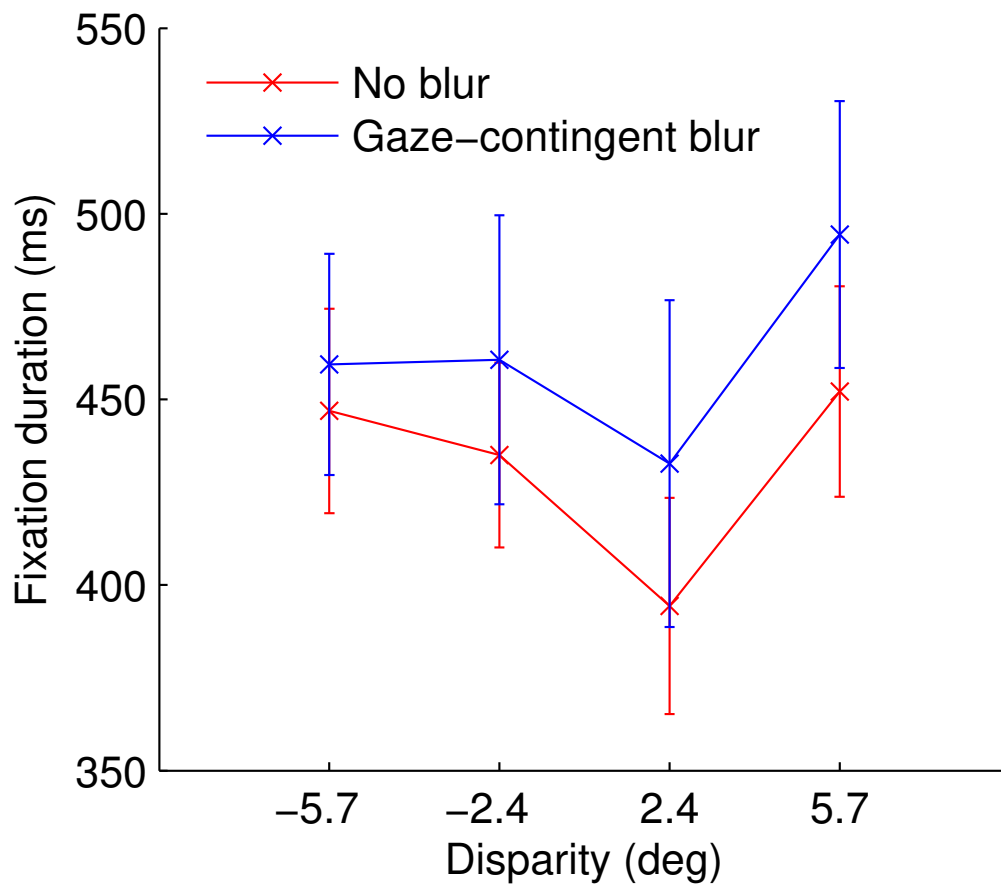


Figure 4.6: **Fixation duration while viewing the stimuli as a function of disparity.** The data are the mean for the 6 observers, error bars represent  $\pm 1$  SEM. The blue curve shows the gaze contingent blur condition, the red curve shows the focused everywhere condition.

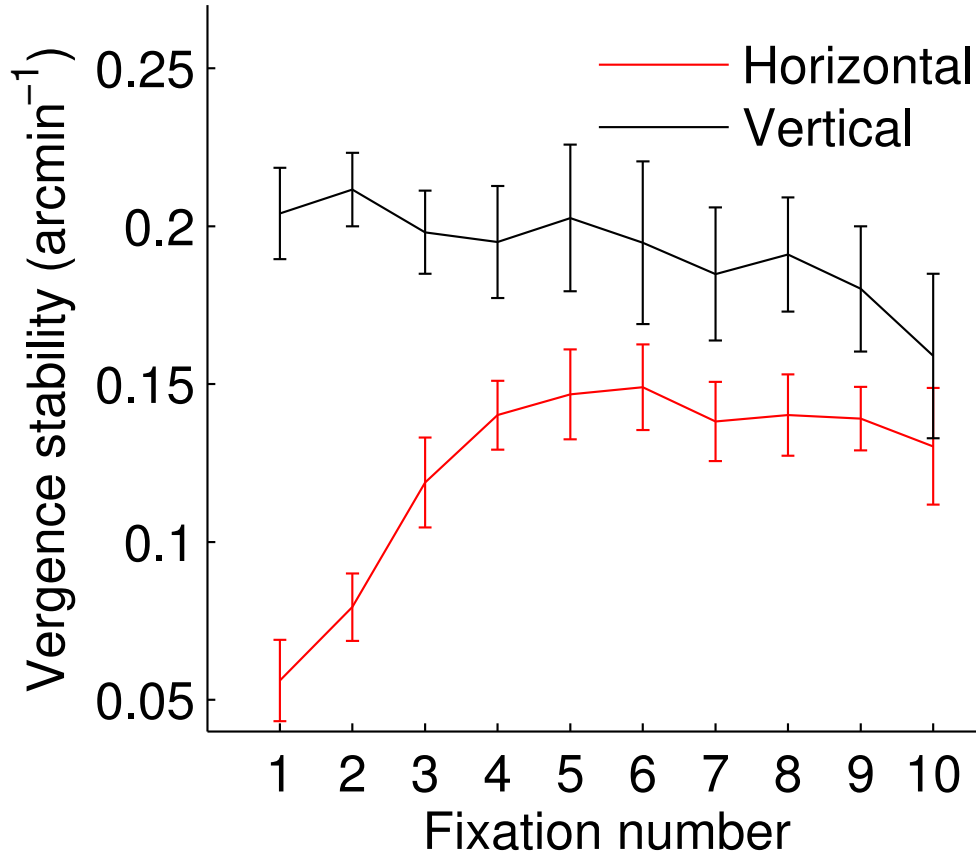


Figure 4.7: **Vergence stability.** Vergence stability for the first 10 fixations (where available) of every trial. The data are the mean of medians for the 6 observers, error bars represent  $\pm 1$  SEM. The red curve shows horizontal vergence, the black curve shows the vertical vergence.

out each trial. I further investigated whether blur and disparity had any effect on horizontal vergence stability. Figure 4.8 shows horizontal vergence stability of the first fixation as a function of disparity for blurred (blue) or sharp (red) images.

There was a significant main effect of disparity  $F(3,15)=5.45$ ,  $p<0.01$ , no significant effect of blur  $F(1,5)=0.01$ ,  $p=0.95$  and no significant interaction between blur and disparity  $F(3,15)=2.76$ ,  $p=0.08$ . Horizontal vergence stability therefore increased from crossed to uncrossed disparities.

## 4.4 Discussion

The time-course of perceptual fusion was significantly affected by depth away from the initial fixation plane, and by the interaction between stereoscopic depth and gaze-contingent blur. Furthermore, there were systematic changes in eye movement behavior for crossed and uncrossed disparity and the presence of optical blur in the images. The number of fixations per trial increased with the magnitude of either crossed or uncrossed disparity. At high levels of uncrossed disparity, viewers required less time and fewer fixations to perceptually fuse the stereoscopic images when informative distributions of retinal blur were present. This benefit was observed only at high uncrossed disparities, which is consistent with larger fusional ranges for crossed disparities than uncrossed disparities (Yeh and Silverstein 1990). Furthermore the largest gains were obtained by people who required longest time to fuse the images. This indicates that a functional gain can be achieved with the inclusion of gaze-contingent blur that varies naturalistically across the visual

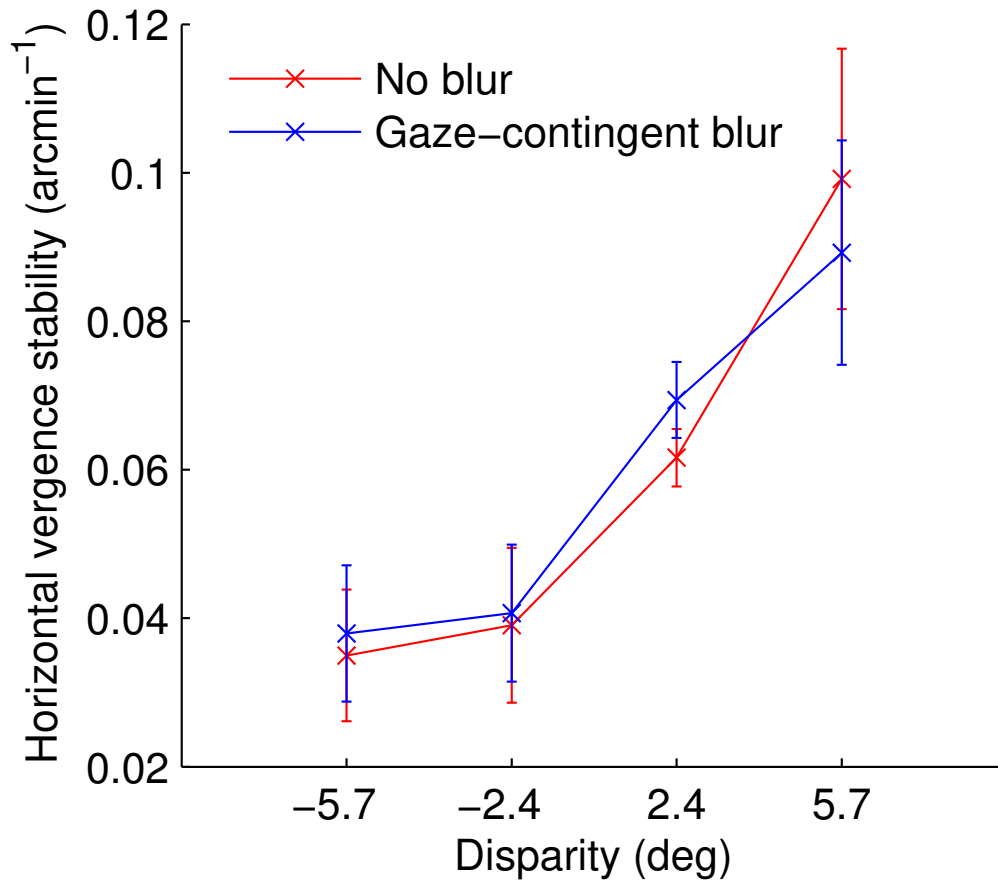


Figure 4.8: **Horizontal vergence stability of the first fixation of every trial as a function of disparity.** The data are the mean of medians for the 6 observers, error bars represent  $\pm 1$  SEM. The blue curve shows the gaze contingent blur condition, the red curve shows the focused everywhere condition.

field and people with greatest difficulty with stereoscopic disparity may benefit most from the addition of blur. My gaze contingent blur implementation necessarily attenuated high spatial frequency information in the visual periphery. However, depending on the level of blur and the retinal eccentricity, this blur may not be detected (Hilz and Cavonius 1974) or may appear to be perceptually sharp (Galvin et al. 1997). My results suggest that blur in the visual periphery, which reduces the contrast of geometric information at high spatial frequencies, does not hinder stereoscopic fusion. Instead, stereoscopic depth and peripheral blur interact and facilitate the perceptual fusion of stereoscopic images. These findings are in good agreement with previous research using stimuli with fixed levels of blur across the stimulus (Siderov and Harwerth 1995; Schor, Wood, and Ogawa 1984b), showing that binocular sensory fusion is greater for low spatial frequency gratings. In the display described in this work, as in natural vision, blur increases with depth disparity in the visual periphery, and this facilitates perceptual fusion of high peripheral disparities.

The presence of peripheral blur also lengthens fixation duration when viewing stereoscopic stimuli, even though the foveal image remained sharp at all times. Previous researchers have demonstrated that stereoscopic images that contain natural distributions of blur are associated with reduced levels of visual fatigue (D. M. Hoffman et al. 2008). Here I show that naturally blurred images also affect eye movement behavior, by reducing the number of fixations, without increasing the total time to fusion. Artificially sharp images elicit more fixations that are each shorter, which may be associated with visual discomfort.

Horizontal vergence stability is greater when viewing uncrossed disparities (objects behind the screen), and worsens when viewing crossed disparities (objects in front of the screen). This is consistent with reports of more visual discomfort when viewing crossed disparities (i.e. stimuli perceptually popping out of the screen) at near distances (Shibata et al. 2011). Informative peripheral blur from my system does not seem to affect this trend.

Natural images have sharp depth discontinuities and owing to micro-saccades during fixation, small changes in eye position could introduce large changes in retinal blur when an observer fixates the edge of an object and gaze drifts between near and far objects in real scenes. When observers fixate depth boundaries in our display, our system may rapidly switch between fixation planes because of eye tracker variance in gaze estimate in addition to fixational micro-saccades. Observers however did not report systematically noticing these rapid switches nor do we notice this effect in natural viewing, and there was no evidence that these switches impacted on observers' experience in the present study.

The gaze contingent display I have developed does not solve the vergence/accommodation conflict when viewing stereoscopic stimuli. Furthermore, I deliberately introduced significant cue conflicts in order to study fusion under these conditions and examine the benefit of including depth-contingent blur in virtual reality applications. I monitored pupil size throughout the experiment, and found that pupil size varied with presented disparity. Since pupil size is linked with accommodative response (Kasthurirangan and Glasser 2005a), it is likely that observers changed accommodation as they changed convergence. Accommodation away from the surface of the display



would have induced additional blur due to the eye’s own optics and this additional source of image blur could have contributed to differences in fusion time. One method to address this issue would be to employ a volumetric display (Akeley et al. 2004; Watt, Akeley, Girshick, et al. 2005) to modulate accommodation as well as blur and disparity.

The blur presented in this gaze-contingent paradigm comes from the optics and image processing of the Lytro camera, which does not perfectly correspond to the blur due to the optics of a human eye. An ecological gaze-contingent model of blur can be envisioned. Assuming an image with infinite depth of field and a calibrated depth map of the visual scene were available as starting points, the mechanism would be as follows. Portions of the scene at the focal distance would be displayed in focus. Portions of the scene away from the focal plane would be displayed filtered through a Gaussian filter with standard deviation equal to the diameter of the blur circle given by Equation 1, or with a Sinc function that introduces periodic phase reversals similar to those in optically blurred images (Murray and Bex 2010). This solution would employ a blur model which is closer to the blur introduced by the optics of the human eye than blur introduced by the optics of the Lytro camera. To implement such a system the gaze-contingent algorithm proposed by (Geisler and Perry 2002) could be used as a starting point.

Multifocal intraocular and contact lenses use diffractive optics or zones with differing refractive power to produce retinal images that are in focus for near and distance vision, thus affecting the distribution of blur across the retina (for reviews see Stein 1990; Leyland and Zinicola 2003; Bellucci 2005). The present results suggest that some observers may experience problems

with binocular fusion or changes in oculomotor behavior when blur cues to depth are removed in this way, which is consistent with reported worsening of stereo acuity with multifocal contact lenses (Ferrer-Blasco and Madrid-Costa 2010). The primary endpoint for multifocal lenses is near and distance acuity, however changes in fusion or oculomotor behavior are not currently assessed, although they may contribute to subjective estimates of visual discomfort on quality of life instruments (Hays et al. 2003).

Although my display implementation does not replicate the optics of any individual observer, one of my research goals was to examine the benefit of standardized implementations of naturalistic blur in virtual reality displays. It is not currently feasible to reproduce the individual aberrations of each observer in a consumer device. Nevertheless, I demonstrate how to approximate some of the changes in the distribution of blur and depth in a simple and inexpensive system that could be implemented in standard electronics. I show that a practical implementation of generalized dioptric blur provides measurable benefits in terms of time to fusion in a stereoscopic display. Additionally, the system I propose enables control over multiple cues to depth which can be carefully manipulated in real images of natural scenes.

## Part 2: Myopia

## Chapter 5

# Retinal Blur from Natural Scenes and Eye Shape

### 5.1 Introduction

The eyes of humans and many animals employ a lens to sharply focus the light coming from a specific object onto the high-resolution fovea. The natural environment contains objects located at many different viewing distances, with consequent varied depth structure. Depth structure can be expressed as diopters,  $1/\text{length}$  in meters, a measure that conveniently represents depth in optical systems. Dioptric topography is thus the depth structure of an environment in dioptric units. Via the process of accommodation, the human eye is able to change the power of the crystalline lens and refocus to depth planes with different dioptric demands. However, only one depth plane may be sharply focused onto the center of the retina at any one point in time. Any depth planes seen with the peripheral retina will therefore have varying

degrees of defocus blur. The amount of defocus blur experienced by the peripheral retina will be dependent on the depth structure of the visual world as well as the curved shape of the retina (Stone and Flitcroft 2004).

Defocus blur is detrimental because it hinders detailed vision. However, since it is related to the depth structure of the world, blur could, in principle, be employed to estimate depth (Burge and Geisler 2011). Yet, as we've seen in chapters 3 and 4, the specific role of blur in depth perception is a matter of scientific debate (Mather 1996; Marshall et al. 1996; Mather 1997; Mather and Smith 2002; Watt, Akeley, Ernst, et al. 2005; Held et al. 2010; Vishwanath and Blaser 2010; Held, Cooper, and Banks 2012; Nefs 2012; Maiello et al. 2014; Langer and Siciliano 2015; Maiello et al. 2015b). Defocus blur also plays a role in emmetropization (Smith III, Li-Fang, and Harwerth 1994; Flitcroft 1998; Smith III and Hung 1999; Hess et al. 2006), the process by which our eyes grow from birth to adulthood. Myopia, commonly known as nearsightedness, is a failure of emmetropization in which the eyes grow excessively large and the accommodation system is no longer able to focus far objects onto the fovea. Myopia is a major public health concern (Morgan, Ohno-Matsui, and Saw 2012), since the prevalence of myopia is increasing (Pan, Ramamurthy, and Saw 2012) and reaching staggering percentages in children, particularly in south-east Asia (L. Lin et al. 2004).

The migration to cities and the consequent increased time spent in indoor environments is associated with this alarming increased myopia prevalence. On the other hand, time spent outdoors is known to be protective against myopia development (Jones et al. 2007; Rose et al. 2008; Dirani et al. 2009), and high light levels outdoors might be important for this protective effect

(Ashby, Ohlendorf, and Schaeffel 2009; Smith III, Hung, and Huang 2012; Ngo et al. 2013). However, given that myopia development is related to visual blur, it could be that the different distributions of depth planes, and therefore blur, in outdoors and indoors environments play a role in the protective effect of spending time outdoors. Furthermore, as myopia progresses, the elongated shape of the eye leads to different amounts of peripheral blur, even when patients are optically corrected to experience a sharp foveal image. It is therefore possible that when individuals with myopia wear the best optical correction for their central vision, they will still have impaired peripheral vision.

Using measurements of natural image statistics, Sprague et al. (2016) have recently shown that the distributions of blur across the visual field differ as a function of the task participants are performing and the environment in which participants are operating. Similarly, in a detailed review on the known factors contributing to myopia aetiology, Flitcroft (2012) employed computer-generated images of indoor and outdoor scenes to show that indoor scenes are more dioptrically varied than the outdoor environment and thus produce larger blur variations in the peripheral retina. We may therefore speculate that differences in eye shape might further alter the statistic of blur in myopic and emmetropic eyes in different natural environments. In this chapter, I employ range image data from indoor and outdoor natural scenes originally collected by Howe and Purves (2002). A range image is a dioptric topography, a map of a real scene containing the three-dimensional location of every point in the scene, acquired using a laser range scanner. I developed a simple geometric model of emmetropic and myopic retinal profiles, based

on magnetic resonance imaging (MRI) data from Atchison et al. (2005). I subsequently employed these eye models to quantify the distribution of naturally-occurring blur across the visual periphery of adults with myopia or emmetropia. Finally, I used these models to verify that these differences in peripheral defocus cause best-corrected myopes to have decreased peripheral visual function with respect to normally-sighted, emmetropic observers. The work presented in this chapter suggests the interesting possibility that the processes controlling eye growth may be attempting to obtain a retinal shape that produces blur distributions that match those experienced by a healthy eye in outdoors environments. If that were the case, spending time indoors might influence these processes, as the geometry of the indoors world and of near work tasks produces very different distributions of blur across the retina (Charman 2011b).

## 5.2 Methods

### Geometric Modeling of the Emmetropic and Myopic Eyes

I implemented a geometric model of emmetropic and myopic retinae, based on MRI data (Atchison et al. 2005). The shape of the retinal surface can be mathematically described by a non-rotationally symmetrical ellipsoid with equation (Figure 5.1):

$$\frac{x^2}{R_x^2} + \frac{y^2}{R_y^2} + \frac{(z - R_z)^2}{R_z^2} = 1 \quad (5.1)$$

where  $R_x$ ,  $R_y$  and  $R_z$  are the semidiameters of the ellipsoid along the  $x$ ,  $y$ , and  $z$  axes. The model eye is placed with its pupil in the origin of the coordinate system and the fovea in  $(0, 0, -2R_z)$ . In emmetropes, the semidiameters were measured by Atchison et al. (2005) to be  $R_x = 11.40 \pm 0.47$  mm,  $R_y = 11.18 \pm 0.50$  mm, and  $R_z = 10.04 \pm 0.49$  mm. In comparison, the semidiameters of the myopic eyes were found to be  $R_x = 11.61 \pm 0.55$  mm,  $R_y = 11.72 \pm 0.50$  mm, and  $R_z = 10.72 \pm 0.59$  mm.

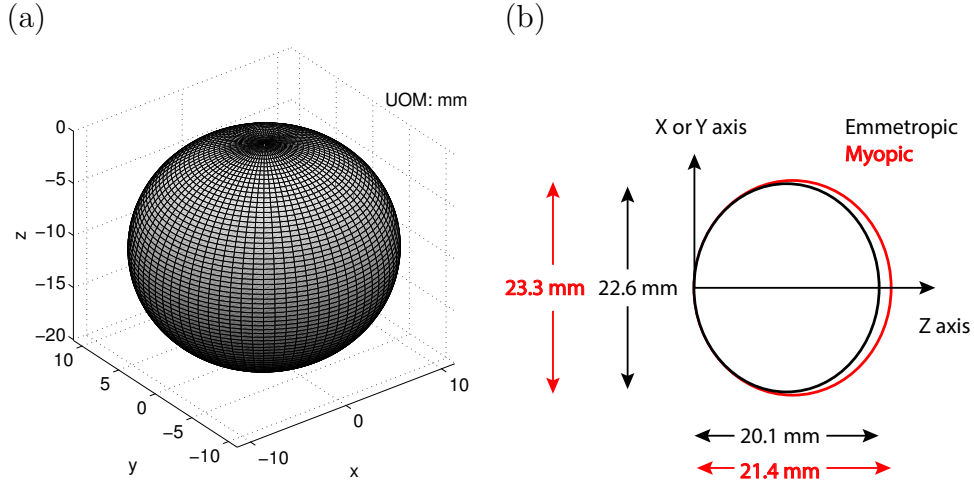


Figure 5.1: **Model of retinal surface.** (a) The retinal surface was modeled as a non-rotationally symmetrical ellipsoid (Eq. 5.1). (b) By slicing through the model retinæ of emmetropes (black) and myopes (red) from published data Atchison et al. (2005), we observe that the retinal shape of both refractive groups is oblate (more width and height than length), but the retina of the average myope is less so, since in myopia eye shape increases more rapidly in the  $z$  dimension than in the  $x$  and  $y$  dimensions.



## Estimating Retinal Blur Distributions

I input range image data of natural scenes (Howe and Purves 2002) to the geometric model retinae to estimate the average amount of blur emmetropes and corrected myopes experience at various locations in the visual periphery when viewing natural scenes. The range image data from Howe and Purves (2002) consists of clouds of 3D points representing the locations of all visible surfaces in a scene. Figure 5.2 shows one such scene where I have color-coded depth. The outdoors database contained 76 outdoors scenes, whereas the indoors database contained 27 indoors scenes. To estimate the average amount of blur at each retinal eccentricity, I perform the following computations.



Figure 5.2: **Example range image.** Depth image of an outdoors scene from (Howe and Purves 2002), where hot colors represent far distances and dark colors represent near distances.

From each scene I select a random point to be the fixated location. The fixation location is drawn from a rectangular uniform distribution 100 degrees in width and 20 degrees in height centered around the initial center of the range map. The distribution of fixations obtained using this method matches well with typical human fixation distributions (Liu, Bovik, and Cormack 2008). The 3D point cloud is then rotated to align this selected location with the fovea of the model eye. This is equivalent to rotating the eye

model to point at the fixated location. Thus the model eye (Figure 5.3) is fixating a point at some known distance  $Z_f$ . Using the thin lens model as approximation of the eye's lens system, the focal length  $f$  of the eye fixating at distance  $Z_f$  can be computed. The thin lens equation for an eye with axial length  $2R_z$  is given by:

$$\frac{1}{2R_z} + \frac{1}{Z_f} = \frac{1}{f} \quad (5.2)$$

Once I have computed the focal length, for every point of coordinates  $P(x_p, y_p, z_p)$  within the eye's field of view I do the following: The line connecting  $P(x_p, y_p, z_p)$  to the center of the eye's pupil  $(0, 0, 0)$  is computed. On that line, I then find the point  $E(x_e, y_e, z_e)$  at which the light coming from  $P$  is focused within the eye using the thin lens equation. Assuming a pupil with diameter  $2pr$ , the uppermost and lowermost borders of the pupil are  $(0, pr, 0)$  and  $(0, -pr, 0)$ . For these calculations I assumed a pupil diameter of 2 mm (Watson and Yellott 2012). I calculate the lines connecting  $(0, pr, 0)$  and  $(0, -pr, 0)$  to  $(x_e, y_e, z_e)$ . I then find the intersections of these two lines with the surface of the retinal ellipsoid, which have coordinates  $(x_{cu}, y_{cu}, z_{cu})$  and  $(x_{cl}, y_{cl}, z_{cl})$ . Finally, the distance between  $(x_{cu}, y_{cu}, z_{cu})$  and  $(x_{cl}, y_{cl}, z_{cl})$  is taken to be the diameter of the blur circle of confusion projected from  $P(x_p, y_p, z_p)$  onto the retina.

By performing these computations for every range image available, I can thus estimate the average amount of blur at any position throughout the visual field of the eye model.

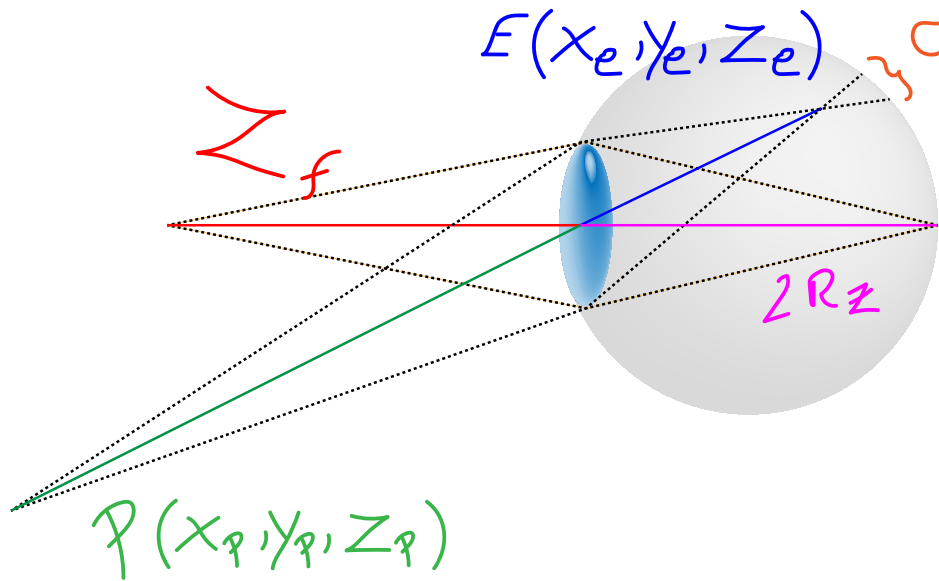


Figure 5.3: **Steps to geometrically estimate retinal blur.** An eye with axial length  $2R_z$  is fixating and focused at a distance  $Z_f$ . A point  $P(x_p, y_p, z_p)$  in the eye's visual field will be in focus within the eye at  $E(x_e, y_e, z_e)$  and will be blurred on the retina over a circular region with diameter  $c$ .

## Measuring Peripheral Contrast Sensitivity in Human Observers

**Participants** Eighteen subjects with normal or corrected to normal vision participated in the study. Of these, 10 were myopes (5 females, mean[range] age: 22[19,30] years, mean[range] refraction: -3.5[-1,-5.25]) and 8 were emmetropes (2 females, mean[range] age: 23[19,31] years). To classify subjects as myopes, subjects who wore glasses or contact lenses were asked to bring their prescription glasses to the lab and the lens power of their spectacles was recorded using a NJC-1 Manual Lensmeter.

**Stimuli** Subjects sat in front of the computer monitor and were required to perform a 26 alternate forced choice letter identification task. Stimuli were 1 octave log cosine band-pass filtered letters presented on a uniformly gray background. Letters were chosen over gratings because they yield the same results (McAnany and Alexander 2006) and increase test efficiency by 1) increasing familiarity for naive observers and 2) reduce guessing rate to 1/26 (Hou et al. 2015).

**Apparatus** The experiment was created using Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) with Matlab version R2014a (Mathworks). Stimuli were presented on a gamma-corrected Dell P2815Q monitor with a mean luminance of 125 cd/m<sup>2</sup>, a resolution of 3840x2160 pixels at 30 Hz, run from an AMD Radeon HD 7000 graphics processing unit. Observers were seated 50cm in front of the monitor with their heads set in a chin rest. Display dot pitch was 0.16 mm and the monitor subtended 70x40 degrees.

**Design and Procedure** It is well known that contrast sensitivity varies lawfully as a function of spatial frequency (Pelli and Bex 2013). I employed an adaptive algorithm that efficiently estimates the whole Contrast Sensitivity Function (CSF) in few trials. The algorithm is described in detail by Lesmes et al. (2010). The CSF at each tested eccentricity was estimated in 50 trials. Each trial, a test letter was selected with spatial frequency and contrast under the control of the adaptive algorithm. The test letter was shown during a 1 second interval in which the target contrast was smoothly ramped on and off over time (Gaussian contrast window with  $\sigma = 250$  ms). Following the stimulus presentation, an observer was shown all 26 possible letter targets and was required to indicate, via mouse click, which of the 26 letters the observer thought was presented.

Prior to the full experiment, subjects were binocularly trained at the task with gaze compliance assured through the use of an EyeTribe eye tracker running at 60 Hz (Dalmaijer 2014). Once subjects were confident they could perform the task maintaining steady fixation, an occluder was placed in front of one of their eyes, and the experiment was performed monocularly to avoid the effects of binocular summation. The tested eye was the most myopic eye in myopic observers, and the non-dominant eye in emmetropic observers. Test stimuli were presented at 5, 10 or 15 degrees in the nasal visual field in order to avoid the physiological blind spot. Thus each subject completed 150 trials.

**Analysis** For each estimated CSF curve from every observer and each condition, both CSF acuity and the AULCSF were computed. The CSF acuity

is defined as the spatial frequency at which Contrast Sensitivity falls to zero. The AULCSF is instead a commonly employed, broad measure of visual function across all spatial frequencies (Applegate, Hilmantel, and Howland 1997; Oshika et al. 2006; Lesmes et al. 2010). For each refractive group, the average CSF acuity and AULCSF at every tested eccentricity was computed. To assess for statistical significance, 95% confidence intervals were computed from 5000 bootstrapped samples. The bootstrapped samples were fit with a Gaussian function, and the mean and standard deviation of fitted distributions were employed to Z-transform group means and compute p-values. At every tested eccentricity, group averages that fell outside the bounds of each other's confidence intervals (corresponding to p-values  $< 0.05$ ) were considered to be significantly different from one another.

## 5.3 Results

### Blur throughout the visual field

I input range image data of indoor and outdoor natural scenes from Howe and Purves (2002) into the model emmetropic and myopic retina to compute the distribution of naturally occurring blurs across the visual field. I define blur as the diameter of the circle of confusion at the plane of the retina for each image point. Figure 5.4 shows the average amount of signed blur throughout the central 40 degrees of the visual field of an emmetropic and myopic eye in both outdoor and indoor scenes. Although blur is an unsigned quantity, for graphical purposes I arbitrarily assigned negative blur values to indicate blur that is due to objects being closer than the fixation distance, while positive

blurs relate to image locations that are farther than the fixation distance. In both outdoor and indoor scenes, blur is  $\approx 0$  near the fovea and increases with eccentricity. In outdoor scenes (Figure 5.4a,c) negative blur due to near objects increases rapidly in the lower visual field, and positive blur due to far objects increases in the upper visual field. In indoor scenes instead (Figure 5.4b,d) blur increases throughout the whole visual field and is primarily due to objects nearer than fixation, in both lower and upper visual field. This is typically due to the prevalence of overhead objects, ceilings and walls in indoor scenes. The pattern of blurs are similar in emmetropic and myopic model eyes, but in the myopic eye model blur grows more rapidly in the visual periphery.

## Blur as a function of visual angle

I quantified the difference in blur throughout the visual periphery of myopes and emmetropes by averaging the unsigned amount of blur at every eccentricity from the fovea of the myopic and emmetropic eye models for outdoors and indoors scenes. Figure 5.5 shows the distribution of blurs at increasing eccentricities away from the fovea of the emmetropic and myopic eyes. Myopic eyes are subjected to greater peripheral blur than emmetropic eyes.

To estimate at what eccentricity the blur distributions differ in the two kinds of retinæ I calculated mean and 95% confidence intervals of blur as a function of retinal eccentricity for myopic and emmetropic eyes. For outdoors and indoors natural scenes separately, I took the eccentricity at which the means of each refractive group fell outside of each other's 95% confidence range to be a conservative estimate of where the average amount of blur is

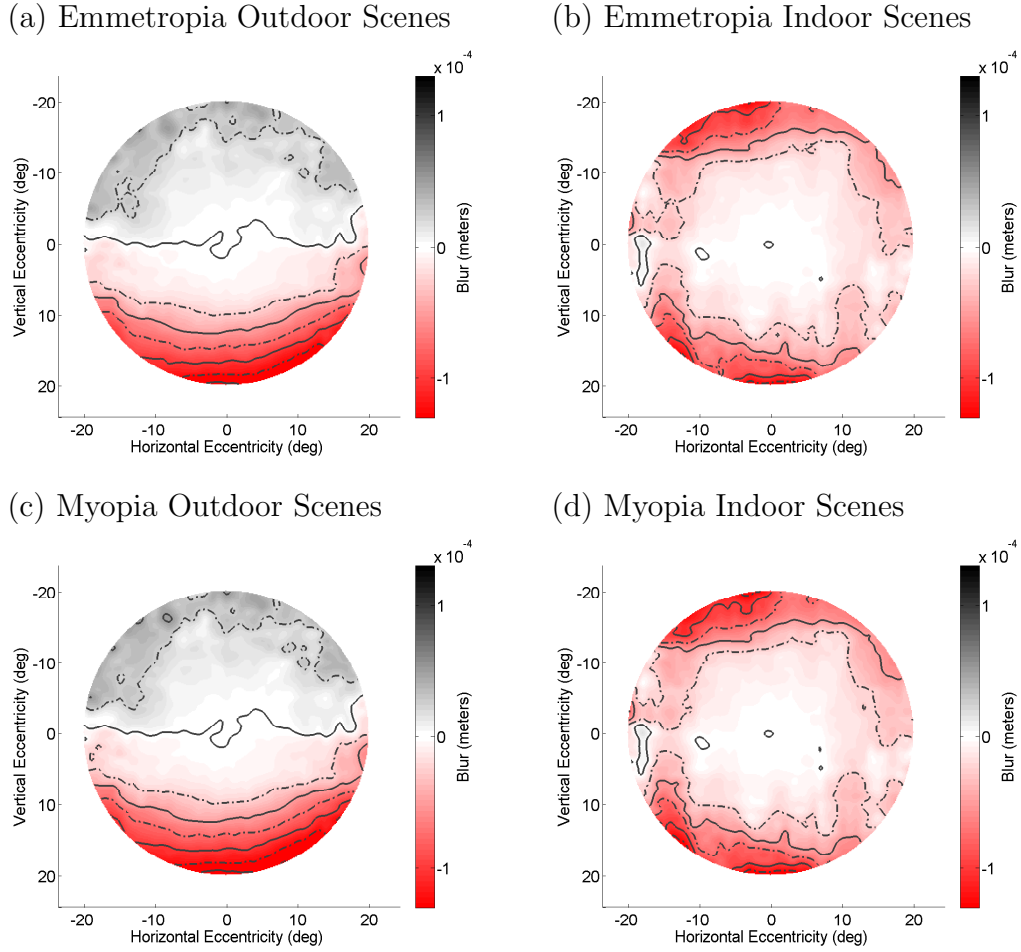


Figure 5.4: **Blur distributions throughout the visual field.** The average signed blur is plotted for each point in the central  $40^\circ$  of the emmetropic (a,b) and myopic (c,d) visual field for both (a,c) outdoor scenes and (b,d) indoor scenes. Gray represents positive blur due to points farther than fixation, and red represents negative blur due to points nearer than fixation. Note the sign of blur is assigned arbitrarily for graphical purposes as blur is a mathematically unsigned quantity. Isocontour black lines are filled for  $0.5 \times 10^{-4}$  meter intervals and dotted for  $0.25 \times 10^{-4}$  meter intervals.



significantly different. For the outdoors scenes, myopic eyes have significantly more blur starting at 9 degrees eccentricity (Figure 5.5a, green asterisk), whereas for the indoors scenes data set, myopic eyes have significantly more blur starting at 14 degrees eccentricity (Figure 5.5b, green circle). Figure 5.5c shows how the distributions of blur for myopes and emmetropes computed from all natural scenes gradually separate with increasing eccentricities.

Thus, at around 10 degrees in the visual periphery of myopic model retinæ the amount of blur departs significantly from the amount of blur in the visual periphery of emmetropic model retinæ. The model predicts that the distribution of blur from natural images does not differ between emmetropic and myopic eyes near the fovea, but myopic peripheral vision should differ regardless of optical correction since more blur is present in the myopic peripheral retina in natural vision.

## **Contrast Sensitivity Falloff in the Visual Periphery of Myopes and Emmetropes**

Because of the differences in eye shape, myopic and emmetropic peripheral vision should differ regardless of optical correction. I proceeded to directly test this prediction in human observers. Specifically, I hypothesized that greater amounts of peripheral blur predict a steeper decline in contrasts sensitivity in myopic observers since blur attenuates high spatial frequency information faster than low.

I tested this hypothesis by measuring the falloff in contrast sensitivity with retinal eccentricity in eight emmetropes and ten best-corrected myopes. The full CSF was assessed at 5, 10 and 15 degrees in the visual periphery

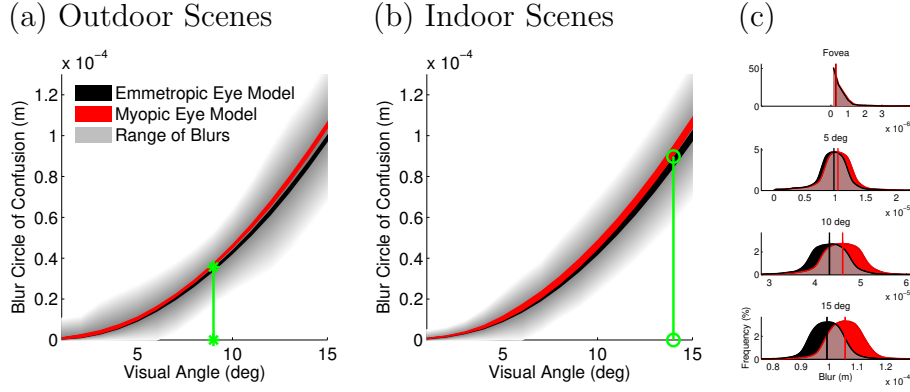


Figure 5.5: **Blur in emmetropic and myopic model eyes.** (a) and (b) show blur as a function of visual angle for outdoor and indoor scenes respectively. Black and red curves are mean level of blur for emmetropes and myopes respectively, the width of the curves represents the 95% confidence range. Gray shaded region represents range of observed blur levels across both eye models. Green markers specify the eccentricities at which the means for each refractive group falls outside of each other's confidence range in indoors (circle) and outdoors (asterisk) scenes. (c) Distributions of blur for myopes (red) and emmetropes (black) at the fovea and at 5, 10 and 15 degrees eccentricity. The distributions are computed from all natural scenes (both indoors and outdoors). Vertical bars are the median level of blur for each group at each eccentricity.

of human observers using a modified adaptive testing procedure (Lesmes et al. 2010). From each measured CSF I then extracted estimates of the CSF Acuity, the spatial frequency at which Contrast Sensitivity falls to zero, and area under the log-CSF (AULCSF), a commonly employed metric (Applegate, Hilmantel, and Howland 1997; Oshika et al. 2006; Lesmes et al. 2010) which provides a broad measure of visual function across all spatial frequencies. I then computed mean CSF acuity, mean AULCSF and corresponding bootstrapped confidence intervals for each group and condition. Group averages for different conditions that fell outside the bounds of each other's confidence intervals were considered to be significantly different from one another.

Figure 5.6a shows CSF acuity as a function of eccentricity for both myopic (red) and emmetropic (black) observers. Acuity decreases as a function of eccentricity for both refractive groups, and is statistically lower in myopes only at 15 degrees ( $p=0.025$ , bootstrapping test). Figure 5.6b shows AULCSF as a function of eccentricity for myopic (red) and emmetropic (black) observers. AULCSF decreases as a function of eccentricity for both refractive groups, and is statistically lower in myopes at 10 ( $p=0.030$ ) and 15 ( $p=0.00017$ ), but not at 5 ( $p=0.082$ ) degrees of eccentricity.

## 5.4 Discussion

In this chapter, I have investigated how blur throughout the visual field is due to both the depth structure of the natural world as well as the shape of the retinal surface. I found that the retinal distribution of blur in outdoor

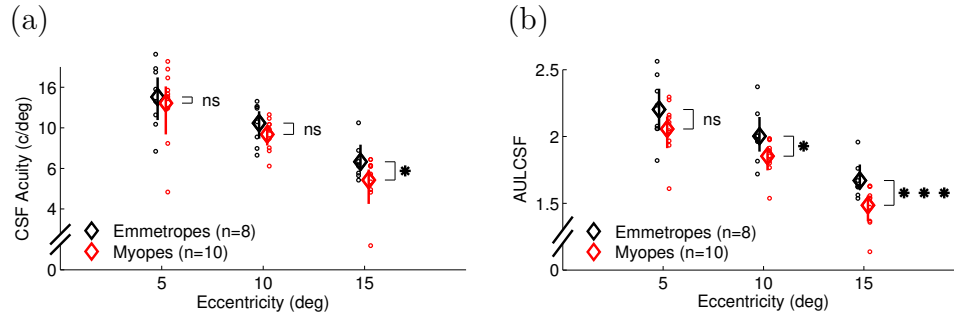


Figure 5.6: **Peripheral visual function in myopes and emmetropes.**

(a) CSF Acuity and (b) AULCSF as a function of eccentricity in myopic (red) and emmetropic (black) observers. Diamonds are means, bars are 95% bootstrapped confidence intervals, circles are individual subject data. Note that y axis in figure (a) is log-scaled. \*  $p < 0.05$ , \*\*\*  $p < 0.001$

scenes is strikingly different from the retinal distribution of blur experienced in indoor scenes. Furthermore, I have shown that best-corrected myopes, who have sharply focused images in the center of their visual field, have greater amounts of peripheral blur than normally-sighted emmetropic subjects. This is due to the shape of myopic eyes, which are larger, and particularly longer, than the eyes of emmetropes. Lastly, I have shown that increased peripheral blur translates to a steeper decrease in visual function in the periphery of myopes compared to emmetropes.

These results are consistent with the notion that the distribution of depths and blurs experienced in outdoors natural scenes may be involved in healthy eye growth. Outdoors, retinal blur in the upper visual field is due to objects being farther than fixation, whereas blur in the lower visual field is due to objects being closer than fixation. Indoors, blur increases in the visual

periphery primarily because of objects being closer than fixation. These findings are in line with those recently reported by Sprague et al. (2016), who analyze the natural statistics of blur within the central 20 degrees of the visual field. Compare the current findings (Figure 5.4) with Figure 4 from Sprague et al. The distribution of blur we observe here from outdoors scenes closely resembles that observed by Sprague et al. (2016) during their “outside walk” activity, whereas the distribution of blur we observe from indoors scenes closely resembles that observed during their “Inside walk” and their “Order coffee” activities, although in the latter activity blur is due primarily to objects being farther than fixation. In their “Make sandwich” activity, similarly to the outdoors results reported here, blur in the upper visual field is due to objects being farther than fixation, whereas blur in the lower visual field is due to objects being closer than fixation. However, in this indoors, near work activity, blur grows more rapidly in the upper than in the lower visual field, a pattern of results opposite to what can be seen for outdoors scenes. This is due to the geometry of blur, and may be understood by inspecting Figure 1B from Vishwanath and Blaser (2010). For near fixation distances (28 cm curve) there is a limited range of blurs due to objects closer than fixation, and a larger range of blurs due to objects farther than fixation. For far fixation distances instead (10 m curve), blur increases rapidly for objects closer than fixation, but asymptotes to small values for objects farther than fixation.

Blur is known to affect eye growth (Smith III, Li-Fang, and Harwerth 1994; Flitcroft 1998; Smith III and Hung 1999; Hess et al. 2006), and time spent outdoors is protective against myopia (Jones et al. 2007; Rose

et al. 2008; Dirani et al. 2009). The differences I've shown between indoors and outdoors blur distributions raise an interesting possibility. The processes driving eye growth may be attempting to obtain a retinal shape that produces blur distributions that match those experienced by a healthy eye in outdoors viewing. Thus, spending much time indoors might confound these processes, as the geometry of the indoors world, and of near work tasks, produces very different blurs across the retina.

These findings also show that the elongated shape of myopic eyes leads to worse peripheral visual function even when foveal vision is corrected. The elongated myopic retina has a significantly greater amount of blur than the emmetropic retina starting at around 10 degrees in the visual periphery. This increase in peripheral blur leads to a measurable decrease in contrast sensitivity because blur attenuates the contrast of mid to high spatial frequency information. These results are also consistent with observed interactions between peripheral contrast sensitivity and attention in myopia (Kerber et al. 2016). It is also possible that decrease in peripheral visual function in myopia might be due to retinal stretching (Chui et al. 2005; Atchison, Schmid, and Pritchard 2006; Coletta and Watson 2006) which could lead to more widely spaced photoreceptors, or even ganglion cell death, and thus a greater undersampling of the peripheral visual field (Thibos, Walsh, and Cheney 1987). Thus, either because of increased blur or because of resolution loss, the elongated myopic retina leads to a worsening of peripheral visual function compared to emmetropia. Peripheral visual impairment also has important functional consequences since the pattern of blur and contrast sensitivity decrease in the visual periphery affects visual processing, attentional selection

and oculomotor planning (Loschky and McConkie 2002; Maiello et al. 2014).

Modeling the differences in eye shape between myopes and emmetropes is only a first step in understanding emmetropization. More complete eye models could incorporate more complex lens systems than the simple thin lens model employed in this work. This would help differentiate the different contributions of the anterior eye structures, cornea, anterior chamber and lens, which are known to differ across refractive groups (Scott and Grosvenor 1993; Grosvenor and Scott 1994; Carney, Mainstone, and Henderson 1997; Grosvenor and Goss 1998). The influence of the different methods employed to correct for myopia (soft and hard contact lenses, glasses, intraocular lenses, LASIK) may also be studied, such as the effects of optical minification produced by spectacles (see Knapp’s Law) (Chui et al. 2005; Coletta and Watson 2006). By implementing non-uniform sampling throughout the retina it may be possible to conclusively differentiate the contributions of optical quality and photoreceptor density (Atchison, Schmid, and Pritchard 2006; Strang, Winn, and Bradley 1998). Finally, by feeding the output of the model retinae into neural models of the first stages of cortical visual processing (Heeger, Simoncelli, and Movshon 1996; Chessa, Maiello, Bex, et al. 2016), it might be possible to investigate how the brain adapts to changes in eye shape, whether neural processing differs across refractive groups, and the role of neural processing in correct eye growth (Maiello, Harrison, et al. 2015; Carroll et al. 2016).

## Chapter 6

# Blur Perception Throughout the Visual Field in Myopia and Emmetropia

### 6.1 Introduction

Blur perception is an elemental feature of the human visual system. A blurred retinal image serves to drive the accommodation and vergence responses, allowing one to see clearly and resolve fine target details (Ciuffreda 1991, 1998; Ciuffreda, Wang, and Vasudevan 2007; Fisher and Ciuffreda 1988; Horwood and Riddell 2009; Z. Lin et al. 2013; López-Gil et al. 2013; Schor 1999). In chapters 3 and 4 we've seen how retinal blur may in some cases also provide information about the range of depths in one's environment (Maiello et al. 2014; Mather 1997; Mather and Smith 2002; Vishwanath and Blaser 2010; Vishwanath 2012; Watt, Akeley, Ernst, et al. 2005) (but see Langer



and Siciliano 2015; Maiello et al. 2015b). As such, blur perception is crucial to one’s ability to navigate the world easily, and to accurately perform daily tasks such as reading and driving (Poulere et al. 2013; Wood et al. 2014). Furthermore, it has been suggested that continuous retinal defocus may be a causative factor in refractive error development, although the evidence is conflicting (Cufflin, Mankowska, and Mallen 2007; Flitcroft 1998; Hess et al. 2006; Norton, Siegwart, and Amedo 2006; Rosenfield and Abraham-Cohen 1999; Schmid et al. 2002; Smith III and Hung 1999; Smith III, Li-Fang, and Harwerth 1994; Strang et al. 2011; Taylor et al. 2009; Vera-Diaz et al. 2015; Wallman and Winawer 2004; Wildsoet and Wallman 1995).

There are two major aspects involved in the perception of blur: blur detection and blur discrimination (Ciuffreda, Wang, and Vasudevan 2007; Wang, Ciuffreda, and Irish 2006). Blur detection refers to the amount of defocus necessary for one to first perceive or notice the presence of blur. Blur discrimination refers to the amount of defocus necessary for one to perceive an already blurry target as just noticeably blurrier. Both aspects contribute to our impression of image quality (Wang and Ciuffreda 2006). Overall, the perception of blur is a complex process that depends on the eye’s optical quality (i.e. aberrations) as well as both retinal and higher level neurophysiology (Ciuffreda, Wang, and Vasudevan 2007; Mather and Smith 2002; Wang and Ciuffreda 2004, 2005a, 2005b). Visual attention, sharpness overconstancy, and target attributes such as luminance, contrast, texture and size also contribute to the perception of image blur (Christman 1990; Galvin et al. 1999; Wang and Ciuffreda 2004, 2005b).

Researchers have studied blur perception within the human fovea exten-

sively (Campbell 1957; Jacobs, Smith, and Chan 1989; Oshima 1958; Rosenfield and Abraham-Cohen 1999; Walsh and Charman 1988), but there is relatively limited information about blur perception throughout the peripheral retina (Ronchi and Molesini 1975; Wang and Ciuffreda 2004, 2005b, 2005a; Wang, Ciuffreda, and Irish 2006), even though peripheral blur strongly impacts peripheral visual function (Maiello, Harrison, et al. 2015; Rosén, Lundström, and Unsbo 2011). Altogether, studies on peripheral blur perception have found that blur detection and discrimination thresholds increase progressively with retinal eccentricity, and that blur detection is initially less sensitive than blur discrimination. However, most of these studies do not measure thresholds directly, rather they infer blur sensitivity by evaluating each subject’s depth-of-focus, either with ophthalmic lenses or by manually displacing the test target, and most have evaluated perception of blur only monocularly (Wang and Ciuffreda 2004, 2005a; Wang, Ciuffreda, and Irish 2006).

To understand how blur is employed by the visual system it is necessary to further study blur perception throughout the peripheral retina. Studying blur perception at different eccentricities may also shed light on emmetropization and its failure, as the interaction between peripheral and central retina may be significant to this process (Huang, Hung, and Smith III 2012; Smith III, Campbell, and Irving 2013; Smith III et al. 2007; Wallman and Winawer 2004; Yamaguchi et al. 2013). Binocularity (the neuronal integration of information from two eyes) needs to also be considered when investigating perception of blur. Previous studies have found that binocularity improves visual performance through probability summation under blur conditions (Banton

and Levi 1991; Heravian, Jenkins, and Douthwaite 1990; Plainis et al. 2011).

In this chapter I assess blur perception in the near peripheral retina, both monocularly and binocularly, in both emmetropic and myopic subjects. I employed naturalistic stimuli (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001; Wallis and Bex 2012) blurred at fixed eccentricities (up to 12 degrees) and measured blur threshold functions employing an adaptive 4AFC paradigm (Vul, Bergsma, and MacLeod 2010). Blur perception in the periphery may vary both because of increasing intrinsic noise (optics and sampling are degraded in the visual periphery) but also because of different neural resources allotted to blur perception throughout the visual field. I differentiate these accounts by estimating both Intrinsic Blur and Blur Sensitivity as a function of eccentricity, refractive status, and monocular or binocular viewing conditions.

## 6.2 Methods

### Subjects

A total of 39 young adult subjects (mean  $\pm$  sd age:  $24.8 \pm 3.8$  years; 19 myopes) were recruited from staff and students of the New England College of Optometry to participate in this study. Following a vision screening that comprised an ocular health evaluation and an ocular history questionnaire, subjects who met all inclusion criteria were enrolled in the study. Criteria for subjects' inclusion were: (1) within 18-32 years of age, (2) best corrected visual acuity (BCVA) 20/20 or better in each eye, (3) refractive error (spherical equivalent, SE) between +0.75 hyperopia and -14.00DS myopia with

$\leq 1.50$ DC of astigmatism or  $\leq 1.00$ D anisometropia, (4) contact lens wearer if myopic refractive correction was needed, (5) no current binocular vision or accommodative dysfunction, (6) not using drugs that may affect their vision, (7) no history of surgery or eye disease that may have resulted in visual consequences, and (8) adequate hearing, language skills and mental ability to understand the consent process and the instructions given during the experiment.

Subjects' refractive error for each eye was determined by objective refraction with an open-field autorefractor (Grand Seiko WR5100K) followed by binocular subjective refraction with binocular balancing and evaluated by the observer's best-corrected visual acuity. Axial length measurements were performed with a Haag-Streit Lenstar LS900 optical biometer. Subjects were grouped based on their refractive error. Myopia ( $n=19$ ) was defined as a SE in each eye between  $-0.50$ DS and  $-11.00$ DS (Mean:  $-5.88 \pm 3.35$ DS). Emmetropia ( $n=20$ ) was defined as SE in each eye between  $-0.25$ DS and  $+0.50$ DS (Mean:  $+0.15 \pm 0.24$ DS).

## Apparatus

The experiment was programmed with the Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) in Matlab. Stimuli were presented on a gamma corrected ROG SWIFT PG278Q Asus monitor with a resolution of  $2560 \times 1440$  pixels (display dot pitch 0.233 mm) running at 120 Hz from an NVidia GeForce GTX 780 graphics processing unit. Subjects were seated 40 cm in front of the monitor with their heads stabilized in a chin and forehead rest. The monitor subtended  $73 \times 46$  degrees of visual angle.

## Stimuli

Stimuli were 46x46 degree patches of dead leaves (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001); examples can be seen in Figure 6.1. Stimuli were constructed from a set of 2000 ellipses each assigned a center position, orientation, aspect ratio, and luminance drawn from pseudo-random uniform distributions. The length of each ellipse semi-axis was randomly selected to be between 0.1 and 10 degrees of visual angle. Each image was divided in four sectors: upper, right, lower, and left. Blur was applied to each sector by Gaussian filtering in the frequency domain by an amount under the control of the adaptive algorithm described below. Three sectors were blurred by the same amount of *pedestal* blur, whereas the fourth was blurred by a greater amount of *pedestal + increment* blur. To measure blur discrimination throughout the selected visual field, a circular central portion of the stimulus up to 0, 4, 8, or 12 degrees of eccentricity was sharply rendered, with no blur (Figure 6.1a-d). Two green concentric circles of 0.3 degrees and 3 degrees respectively were presented in the middle of each stimulus image to serve as target for central fixation.

## Design

It is well known that the ability of human observers to discriminate differences in blur varies lawfully as a function of the amount of reference blur, and that this blur discrimination function is dipper shaped (Hamerly and Dvorak 1981; Pääkkönen and Morgan 1994; Watt and Morgan 1983, 1984), for a review see (Watson and Ahumada 2011). Thus, blur perception is typically

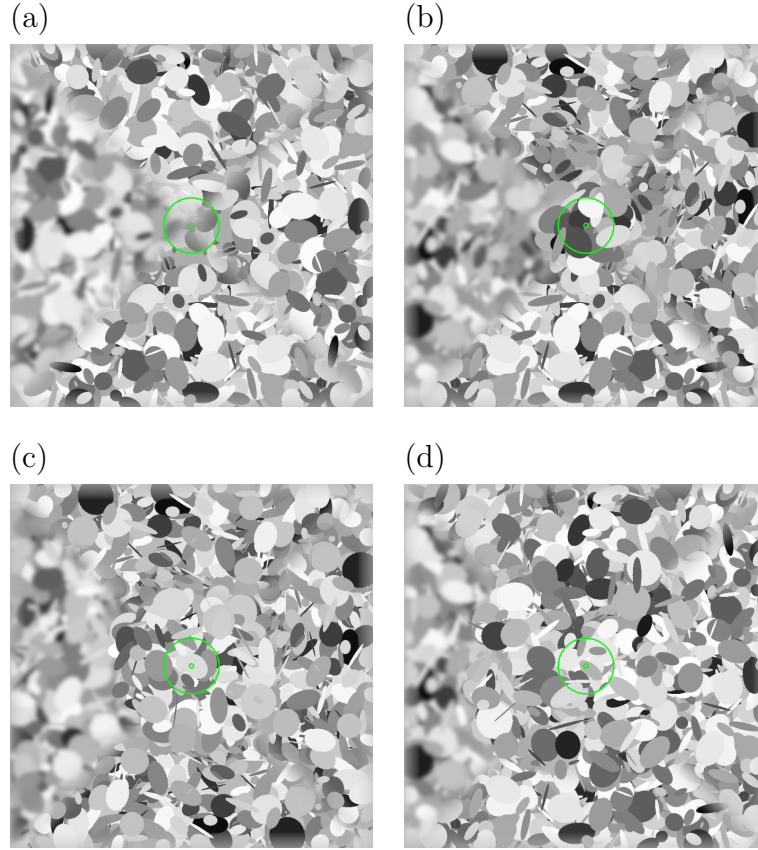


Figure 6.1: **Example stimuli for blur discrimination.** Full field of dead leaves stimuli employed in the study for each eccentricity condition (a-d). In all four example stimuli, the top, right and bottom quadrants are blurred with 0.5 arcmin of *pedestal* blur, whereas the left quadrant is blurred by a greater amount of  $0.5 + 6$  arcmin of *pedestal + increment* blur. In the (a) 0 degree condition, blur is present throughout the visual field. In the eccentric testing conditions (b-d), the dead leaves within the central 4, 8 or 12 degrees were sharply rendered, and blur was present only in the visual periphery.

investigated by measuring blur discrimination thresholds at multiple levels of pedestal blur, and then by fitting the threshold data with a dipper shaped function. Instead of measuring individual thresholds, I employed an adaptive testing procedure (Vul, Bergsma, and MacLeod 2010) to estimate blur dipper functions at 0, 4, 8 and 12 degrees in the visual field both monocularly and binocularly.

I adopted a parameterization of the blur dipper function (Mather and Smith 2002; Murray and Bex 2010) with equation:

$$\Delta b = \sqrt{(1 + 1/S)(\sigma_i^2 + \sigma_e^2)} - \sigma_e \quad (6.1)$$

where blur increment threshold  $\Delta b$  varies as a function of the external (pedestal) blur  $\sigma_e$  applied to the stimulus and is modulated by two parameters: Blur Sensitivity  $S$  and Intrinsic Blur  $\sigma_i$ . The model’s Intrinsic Blur parameter takes into account that a human observer’s visual system has a baseline level of intrinsic blur arising from all optical and neurological sources. The Blur Sensitivity parameter takes into account the proportional increase of the blur threshold with increasing external blur: an observer with high sensitivity will have globally lower thresholds and will thus be overall better at discriminating blur.

The Functional Adaptive Sequential Testing (FAST) (Vul, Bergsma, and MacLeod 2010) algorithm was employed to obtain dipper function estimates. An example of a dipper function estimated for one representative observer in one experimental condition is shown in Figure 6.2. Specifically, for each dipper function estimate the algorithm selected 10 trials at each of 5 pedestal blur levels [0.125, 0.5, 2, 8, 32] arcmin. At these reference blur levels, each

trial the FAST algorithm selected test blur levels that maximized information gain, efficiently constraining the parameter estimates of the underlying dipper function.

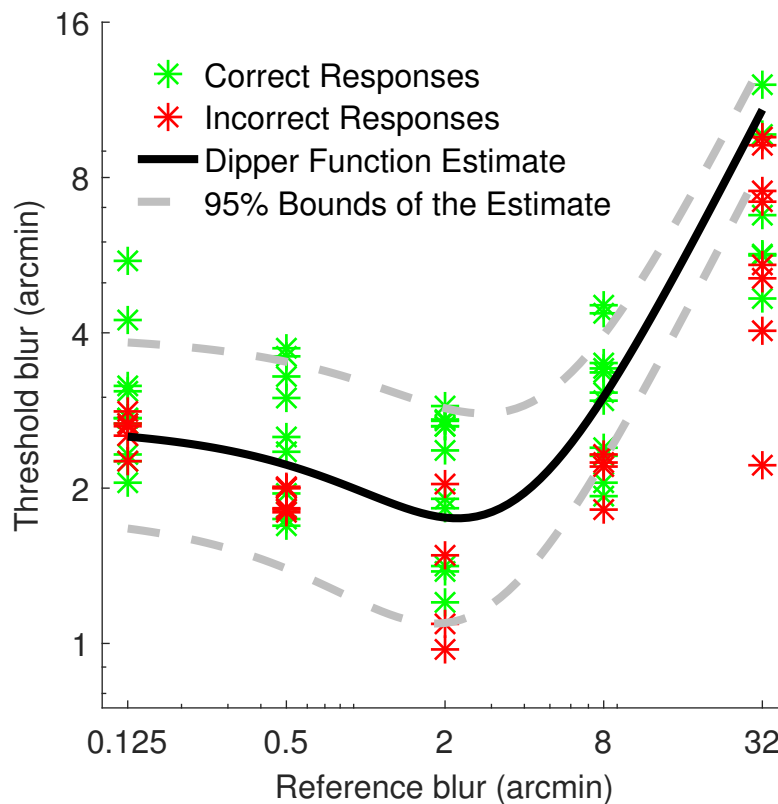


Figure 6.2: **Example of a dipper function.** Example of a dipper function estimated for one representative observer in the binocular, 0 degree eccentricity condition. Asterisks represent individual trials: green and red are correct and incorrect responses respectively. Black curve is the estimated dipper function describing how the observer's threshold blur increment varies as a function of reference blur. Dotted lines are the 95% bounds of the estimated curve.



## Procedure

Subjects were tested both binocularly and monocularly (with their non-dominant eye occluded) in two separate sessions. The session order was randomized across participants. Each session, subjects completed 10 trials at each of 5 pedestal blur levels for each of the 4 eccentricity conditions. Thus each subject completed 200 trials per sessions. When necessary, subjects were corrected with soft contact lenses.

Each trial, subjects were required to fixate the central fixation target. Stimuli were shown for 250 ms, which was too brief for subjects to make eye movements. The 4AFC task was to identify, via button press on the keyboard placed in front of them, which of the four image sections (upper, left, lower or right) was most blurred. Visual feedback was provided by at the fixation mark, which was green following a correct response, or red following an incorrect response.

## Data Analyses

On some conditions and for certain subjects, the FAST algorithm failed to converge. Thus, outlier removal was performed by employing the following procedure. For each estimated parameter, the size of the 95% confidence region of the estimate was computed. For Intrinsic Blur ( $\sigma_i$ ) and Blur Sensitivity ( $S$ ) parameters separately, the upper 90th percentile of the size of the confidence regions was then computed. Finally, data was excluded from those dipper curves in which at least one of the parameter estimates confidence region was greater than this 90th percentile. This procedure thus

excluded parameters estimated with high uncertainty (13% of dipper curves were excluded overall).

Estimated Intrinsic Blur ( $\sigma_i$ ) and Blur Sensitivity ( $S$ ) parameters were analyzed with a 2 (refractive status, between subjects factor) x 2 (viewing condition, within subjects factor) x 4 (eccentricity, within subjects factor) mixed design ANOVA. Blur Sensitivity data did not respect the ANOVA assumption on the normality of the residuals. Thus Blur Sensitivity data was square-root transformed, and ANOVA was re-run on these transformed data which conformed to the ANOVA assumptions on the normality of the residuals. ANOVA results on the original and transformed data did not substantially differ; ANOVA results on the transformed data are reported for statistical rigor. Post-hoc comparisons were conducted via bonferroni corrected t-tests.

## 6.3 Results

Figure 6.3 shows group data dipper functions estimated under monocular and binocular testing conditions at each tested eccentricity. In both monocular and binocular viewing, blur thresholds increase with eccentricity. Also blur discrimination thresholds are generally lower in the binocular testing condition.

Figure 6.4 shows the Intrinsic Blur parameter of the estimated dipper curves. Intrinsic blur is larger under monocular than binocular conditions ( $F(1,42)=12.75$ ,  $p=0.00091$ ) and increases with eccentricity ( $F(3,108)=32.73$ ,  $p=10^{-14}$ ). There was no significant effect of refractive group ( $F(1,38)=0.54$ ,

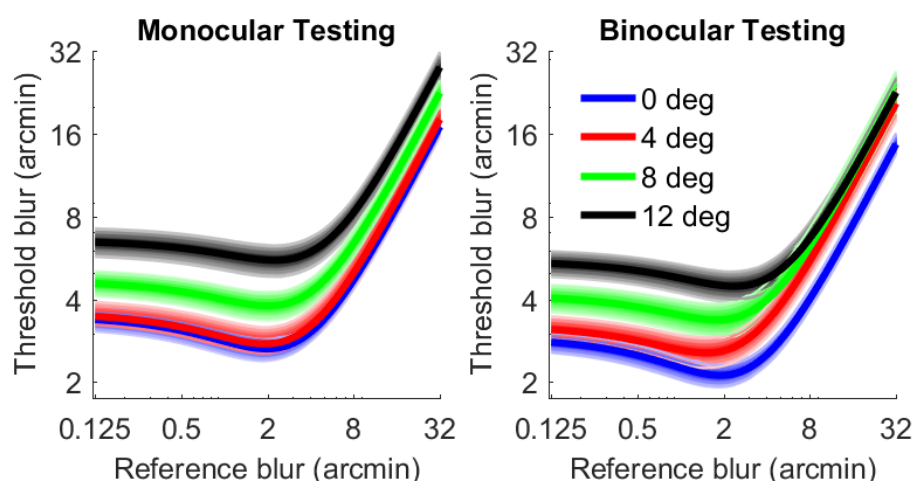


Figure 6.3: **Average blur dipper functions.** Blur dipper functions under monocular (left) and binocular (right) conditions. Curves are averaged across all subjects at each eccentricity (0 degree blue, 4 degree red, 8 degree green, and 12 degree black). Shaded regions are 95% bootstrapped confidence intervals of the mean.

$p=0.47$ ) and no significant two or three way interactions (all  $p>0.6$ ).

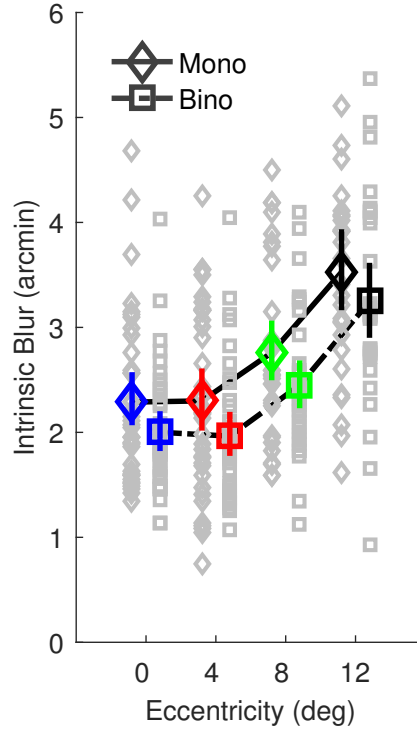


Figure 6.4: **Intrinsic Blur.** Intrinsic Blur as a function of eccentricity in monocular and binocular viewing. Data are averaged across subjects for both monocular (diamonds) and binocular (squares) viewing conditions. Color-coding of eccentricity conditions is as in Figure 7.3. Error bars are 95% bootstrapped confidence intervals of the mean. Gray data points are individual subject data. Monocular and binocular data are shifted left and right for graphical purposes.

Figure 6.5a shows Blur Sensitivity under monocular and binocular viewing. Blur Sensitivity decreases with eccentricity ( $F(3,109)=11.6$ ,  $p=10^{-6}$ ). Comparing monocular and binocular viewing, there was a significant main

effect of Viewing Condition ( $F(1,43)=5.88$ ,  $p=0.020$ ) and a significant two way interaction between Viewing Condition and Eccentricity ( $F(3,75)=4.91$ ,  $p=0.0036$ ). Post-hoc test following the significant interaction between Viewing Condition and Eccentricity revealed that Blur Sensitivity is higher under binocular viewing conditions at the fovea ( $p=0.012$ ). However, there was no significant difference in Blur Sensitivity between binocular and monocular viewing conditions at 4, 8, and 12 degrees of eccentricity (all  $p>0.25$ ). To summarize these results, in Figure 6.5b Binocular Sensitivity Advantage is plotted as a function of eccentricity. That is, I plot the difference between Blur Sensitivity measured binocularly and monocularly at each tested eccentricity. If binocular viewing boosts Blur Sensitivity, we would expect the Binocular Sensitivity Advantage to be positive. We can see that on average, Binocular Sensitivity Advantage was positive, but significantly so only at the fovea.

There was no significant main effect of refractive group on Blur Sensitivity ( $F(1,38)=3.49$ ,  $p=0.069$ ), but a significant two way interaction between Viewing Condition and Refractive Status ( $F(1,39)=4.11$ ,  $p=0.049$ ). Bonferroni corrected independent samples t-tests revealed that Myopes had significantly worse Blur Sensitivity than Emmetropes monocularly ( $p=0.037$ ) but not binocularly ( $p=0.66$ ), as shown in Figure 6.5c.

Lastly, there was no significant two way interaction between Eccentricity and Refractive Status ( $F(3,109)=0.28$ ,  $p=0.84$ ), and no significant three way interaction ( $F(3,75)=1.09$ ,  $p=0.36$ ).

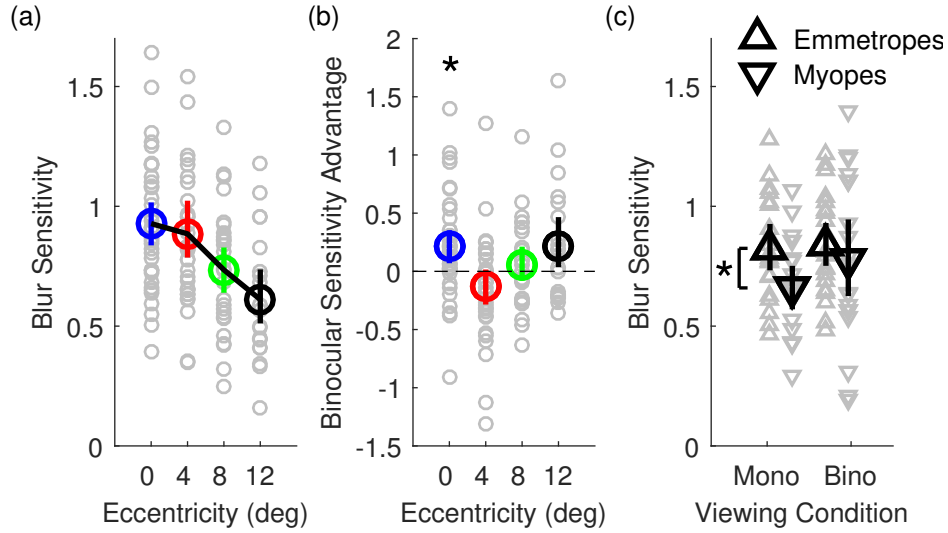


Figure 6.5: **Blur Sensitivity.** (a) Blur Sensitivity as a function of eccentricity averaged across subjects and viewing conditions. (b) Blur Sensitivity Advantage for binocular viewing compared to monocular viewing as a function of eccentricity averaged across all subjects. Color-coding of eccentricity conditions in (a) and (b) is as in Figure 7.3. (c) Monocular and Binocular Blur Sensitivity in Emmetropes (upwards pointing triangles) and Myopes (downwards pointing triangles). Data are averaged across eccentricities and across subjects in each refractive group. Error bars are 95% bootstrapped confidence intervals of the mean. Gray data points are individual subject data. \*  $p < 0.05$

## 6.4 Discussion

In this chapter I have assessed monocular and binocular blur perception in the near peripheral retina in emmetropic and myopic subjects.

It is known that the tolerance of blur increases in the far (Ronchi and Molesini 1975) and near (Wang and Ciuffreda 2004, 2005a) peripheral visual field. Blur perception is also known to follow a typical dipper shaped curve (Watson and Ahumada 2011), with blur discrimination thresholds being lower than detection thresholds for small blur pedestals, and subsequently increasing with increasing levels of reference blur. This pattern is known to be maintained in the visual periphery, where blur discrimination thresholds with detectable levels of blur are lower than blur detection thresholds (Wang, Ciuffreda, and Irish 2006). I have replicated these findings by showing that blur perception worsens in the visual periphery and that performance can be described by a dipper shaped function.

I have extended these findings to show that the increase in blur discrimination thresholds with increasing eccentricity can be attributed to two sources: (1) observers' increased Intrinsic Blur and (2) decreased Blur Sensitivity. This suggests that blur perception worsens in the visual periphery due to poor optical quality and degraded sampling (which leads to an increase in Intrinsic Blur), but also because fewer neural resources are allotted to blur perception in the peripheral visual field (which yields a decrease in Blur Sensitivity).

Plainis et al. (2011) have shown that with visual blur induced by positive lenses, binocularity improved visual acuity and enhanced the P100 component of Visual Evoked Potentials. I extend these findings by showing that

binocularity also improves visual blur perception. These results show that combining the information from the two eyes leads to a reduction in the system's Internal Blur, and also that there is a small but statistically significant advantage in Blur Sensitivity when viewing stimuli binocularly in the fovea. This suggests there may be additional binocular processing of blur in the center of the visual field.

Evidence for differences in the perception of blur across refractive groups is conflicting. Rosenfield and Abraham-Cohen (1999) have found that thresholds for blur detection are impaired in myopes, whereas Cufflin, Mankowska, and Mallen (2007) have found that early onset myopes adapt more strongly to optical blur, yet Schmid et al. (2002) found no statistically significant differences between myopic and emmetropic children in the ability to detect blur. In the current study I have found no differences in the amount of Intrinsic Blur between myopic and emmetropic subjects. I have found instead that myopes have, on average, worse Blur Sensitivity than emmetropes monocularly but not binocularly. This suggests that myopes may process blur less efficiently than emmetropes monocularly but not binocularly. Future investigations should include hyperopic subjects as well as progressing myopes to further study the relationship between blur sensitivity and refractive error development. In the current study blur perception was assessed using rendered Gaussian blur. It is possible that blur perception may differ with different kinds of blur, and that the specific type of defocus blur arising from an observer's own optics may have a role in emmetropization. Future investigations should thus also assess the differences in blur perception with different kinds of blur, such as with testing rigs capable of producing defocus



blur through an observer's own optics (Akeley et al. 2004; Love et al. 2009; Sebastian, Burge, and Geisler 2015) or with sync blur, which contains phase reversals typical of the modulation transfer function of an optical system with a circular aperture such as the human pupil (Murray and Bex 2010)).

# Chapter 7

## Vergence Driven Accommodation with Simulated Disparity in Myopia and Emmetropia

### 7.1 Introduction

The human visual system undergoes development into adulthood. Newborns have very poor visual resolution (Atkinson, Braddick, and Moar 1977; Banks and Salapatek 1978), oculomotor control (Hainline et al. 1984; Roucoux, Culee, and Roucoux 1983) and binocular coordination (Hainline and Riddell 1995). In the first few months of life infants learn to accommodate in order to refocus the retinal image following a change in fixation distance (Banks 1980; Haynes, White, and Held 1965). Binocular oculomotor coordination

begins to develop in the first 6 months of life and is refined into adolescence (Aslin 1977; Fioravanti et al. 1995; Yang and Kapoula 2003). This complex sequence of developmental events can be disrupted and may lead to, amongst other outcomes, incorrect eye growth and thus the development of refractive errors, particularly myopia. Failure of emmetropization (correct eye growth towards a normal end point of emmetropia) may arise due to a multitude of factors that combine a predisposing genetic susceptibility with a major role of environmental factors, such as higher education level. These factors are related to an increased amount of near work (Saw et al. 2002) or decreased time spent outdoors (Rose et al. 2008; Dirani et al. 2009; Deng, Gwiazda, and Thorn 2010; Dirani et al. 2010), for reviews see (Jones et al. 2007; Saw 2003; Charman 2011a; French et al. 2013; Goldschmidt and Jacobsen 2014).

The key to understanding the initiation and progression of myopia requires an understanding of two mechanisms: the environmental factors that disrupt the normal visual feedback loop that controls emmetropization in children and the biological mechanisms that cause a remodeling and growth of the posterior sclera. In this chapter I address the first of these two mechanisms. From studies on animal models we know that disrupting the visual feedback loop through pattern deprivation causes axial elongation to grow uncontrolled, usually resulting in excessive elongation and myopia (Sherman, Norton, and Casagrande 1977; Wiesel and Raviola 1977; Wallman, Turkel, and Trachtman 1978; Troilo and Judge 1993). Biasing the visual feedback with lenses leads the infant animal's eye to grow toward an amount of myopia or hyperopia of the same power as the lens used (for a review see Wallman and Winawer 2004). In humans, the use of plus lenses or lower negative lenses

to prevent or correct myopia does not avoid progression of myopia (Morgan, Ohno-Matsui, and Saw 2012; Walline et al. 2011). However, there is a second mechanism that demonstrates human eye’s ability to respond to changes in retinal image quality. This is ocular accommodation, an alternative measure for investigating how we are able to alter focus in response to retinal image changes.

It has been suggested that myopic children (Gwiazda et al. 1993) and adults (Abbott, Schmid, and Strang 1998; Bullimore, Gilmartin, and Royston 1992; Schmid and Strang 2015) have a decreased ability to accurately accommodate, e.g., larger lags of accommodation and larger variability in accommodation responses compared to emmetropes (Allen et al. 2009; Harb, Thorn, and Troilo 2006; Langaas et al. 2008; Pandian et al. 2006; Price et al. 2013; Strang et al. 2011). Animal studies have also shown an association between inaccurate accommodation and myopia (Troilo, Quinn, and Baker 2007; Troilo, Totonelly, and Harb 2009; Wildsoet et al. 1993). Such inaccuracies of accommodation have been suggested as a causative factor for myopia, as they may lead to blur on the retina and that may prevent normal emmetropization (Goss 1991; Gwiazda, Thorn, and Held 2005; Gwiazda et al. 2004; Hampson, Cufflin, and Mallen 2013; McBrient and Millodot 1986; Millodot 2015). Indeed, blur and depth sensitivity in the visual periphery may play a role in accommodation and emmetropization (for reviews see Charman and Hema 2010; Smith III 2013. Causality may be reversed however, as decreased sensitivity to retinal blur in myopes may instead be the source of accommodation inaccuracies (Cufflin, Mankowska, and Mallen 2007; Rosenfield and Abraham-Cohen 1999). Not all studies have found an

association between myopia and accommodative inaccuracies (Berntsen et al. 2011; Mutti et al. 2007; Rosenfield, Desai, Portello, et al. 2002; Seidel, Gray, and Heron 2005; Taylor et al. 2009).

In this chapter I examine the ability of retinal image quality change in the peripheral visual field to alter accommodation. This approach is a result of exciting animal research in myopia in which it is shown that defocus restricted to half the visual field causes axial elongation for only that half of the retina (Wallman, Turkel, and Trachtman 1978; Wallman et al. 1987). In addition, animals models show that eyes that had laser induced foveal ablation still become myopic in response to minus lenses (Smith III et al. 2007) and that this myopic response is as great as in animals with intact foveae (Huang, Hung, and Smith III 2011). Thus, the peripheral retina can clearly drive axial elongation and the development of myopia. With this in mind, manufacturers around the world are marketing contact lenses to reduce myopia progression. Although their approach is based on solid evidence of the importance of peripheral vision in normal emmetropization, the success of these products has been very limited.

In order to understand how lenses that affect peripheral vision focus may in turn affect emmetropization, it is important to determine how changes in the image may affect the peripheral retina response and elicit accommodative responses. In this chapter I therefore measure accommodation responses to naturalistic stimuli (“dead leaves” images) spanning the visual periphery.

In normal binocular vision, vergence and accommodation responses are tightly associated (Fincham and Walton 1957). Vergence eye movements require opposite rotations of the two eyes that redirect the binocular gaze point

to the depth of the visual target currently of interest. Vergence eye movements are elicited primarily from a combination of retinal disparity signals and accommodation demand. At the same time, accommodation is mainly driven by retinal defocus blur and vergence demand. Thus, in normal vision, vergence and accommodation are closely coupled. One situation in which vergence and accommodation may become uncoupled is when viewing 3D virtual reality, where depth, blur and disparity are usually independent (D. M. Hoffman et al. 2008; Rushton and Riddell 1999; Wann, Rushton, and Mon-Williams 1995). In virtual reality displays, disparity information can drive vergence away from the surface of a 3D monitor, however, since light originates from the surface of the monitor, focus cues will drive accommodation toward the plane of the monitor, creating a cue conflict.

Virtual Reality (VR) technology is in continuous advancement, with recent hype focused on head mounted displays such as the Oculus Rift (see Chapter 12 and Chessa, Maiello, Borsari, et al. 2016). Given the complex interplay between vergence and depth and focus cues, it is fundamental to understand the influence of simulated depth information on accommodative responses as virtual reality systems become increasingly pervasive, particularly for young users whose visual systems may still be in development. One tantalizing idea is the possibility of employing VR technology to manipulate vergence and accommodation stimuli in order to prevent refractive error development or treat its progression (Maiello, Bex, and Vera-Diaz 2014).

In this chapter I investigate the vergence-accommodation link with simulated depth cues in a stereoscopic 3D system in myopes and emmetropes. I assess how the vergence-accommodation link varies with focus, disparity and

pictorial depth cues provided by geometric perspective.

## 7.2 Methods

### Subjects

A total of 25 young adult subjects (mean  $\pm$  sd age:  $24 \pm 2$  years) were recruited for this study. Following a vision screening that comprised ocular health evaluation and ocular history questionnaire, 21 subjects who met the inclusion criteria were enrolled in the study. Criteria for subjects' inclusion were: (1) no history of surgery or eye disease that may have resulted in visual consequences, (2) within 18-32 years of age, (3) best corrected visual acuity (BCVA) 20/20 or better in each eye, (4) not using drugs that may affect their vision, (5) no current binocular vision or accommodative dysfunction, (6) contact lens wearer if myopic refractive correction was needed, (7) refractive error between +1 hyperopia and -14.00DS myopia with  $\leq 1.50$ DC of astigmatism or  $\leq 1.00$ D anisometropia. One emmetropic subject who met inclusion criteria was nonetheless excluded from this study because the photorefractor failed to produce measurements of the subject's accommodative status.

Subjects' refractive error (spherical equivalent, SE) was determined by binocular subjective refraction with binocular balancing and evaluation of the observer's Snellen visual acuity. Objective refraction was measured with a Grand Seiko WR=5100K autorefractor and axial length measurements were performed with a Zeiss IOL Master optical biometer. The refractive error of the observers included in the study ranged from +0.625D to -7.625D (mean

$\pm$  sd SE:  $-1.5 \pm 2.3\text{D}$ ). Of these, 9 subjects were myopes (mean  $\pm$  sd SE:  $-3.6 \pm 2.0\text{D}$ ) and 12 were emmetropes (mean  $\pm$  sd SE:  $+0.1 \pm 0.4\text{D}$ ). Subjects were classified into the two refractive groups as follows: emmetropes were defined as having a SE in each eye between  $+1$  to  $-1\text{D}$ ; and myopes were those subjects with SE in each eye between  $-1$  to  $-14.00\text{D}$  (SE). All subjects were naïve as to the purpose of the experiment.

## Apparatus

The stimuli and experimental codes were programmed with the Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) in Matlab. Stimuli were presented on a gamma-corrected Acer GD235HZ 24" monitor with a resolution of 1920x1080 pixels (display dot pitch 0.2715 mm) running at 120 Hz from an NVidia Quadro FX580 graphics processing unit.

Subjects wore active stereoscopic shutter-glasses (NVIDIA 3D Vision) throughout the experiment. Figure 7.1 schematizes the geometry of the experimental setup. Observers' eyes were at 40 cm (2.5 diopters) from the monitor with the head stabilized in a chin and forehead rest. Thus, the monitor subtended 66x40 degrees of visual angle.

Observers' vergence and accommodative status were recorded at 25 Hz employing a PowerRefractor, an infrared photorefractor (Choi et al. 2000; Schaeffel, Wilhelm, and Zrenner 1993), positioned 1 meter from the observer. The PowerRefractor delivers infrared light into a subject's pupils and simultaneously records the reflection of the infrared light returned through the pupil from the back of the eye. From the pupil light distribution profile, the PowerRefractor determines instantaneous estimates of the refractive state of



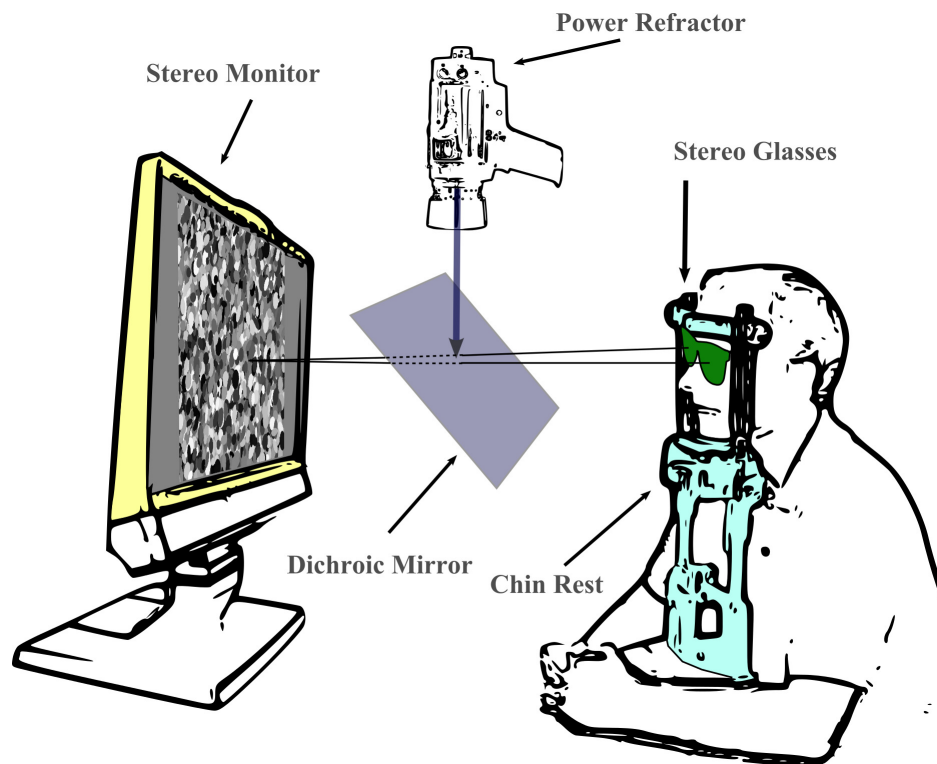


Figure 7.1: **Schematization of the experimental setup to measure vergence and accommodation.** Observers wore NVidia 3D vision shutter glasses and viewed the 3D monitor through a 45 degree tilted dichroic mirror. The Power Refractor camera was placed above the mirror and viewed the observers' eyes reflected in the mirror.

the subject's eyes (Roorda, Campbell, and Bobier 1997; Schaeffel, Wilhelm, and Zrenner 1993), an indirect measure of accommodation. Vergence is computed as the difference in horizontal gaze position between the two eyes, with gaze position for each eye estimated using the relative displacement of the first Purkinje image from the center of the image of the pupil (Brodie 1987; Riddell, Hainline, and Abramov 1994).

Because the PowerRefractor must have a straight ahead view of the observer's eyes, a dichroic mirror was positioned between the observers' eyes and the stimulus display, which allowed light from the monitor to be transmitted to the observers' eyes. At the same time, the mirror reflected infrared light from the PowerRefractor to the observers' eyes, therefore allowing for continuous recording of the observers' vergence and refractive state. The PowerRefractor was run from a dedicated pc.

## Stimuli

### **Step 1: Vergence driven Accommodation to Simulated Disparity**

Stimuli to drive vergence without focus cues for accommodation were two green fixation dots (0.5 degree diameter each) rendered with -3.5 degrees of uncrossed disparity, as illustrated in Figure 7.2a. Each fixation dot was Gaussian filtered with a standard deviation of 5 pixels (12 arcmin) so as to contain no mid or high spatial frequency edge information and thus did not provide a reliable stimulus for accommodation at the surface of the monitor. Given that the spatial structure of the Gaussian fixation dot was constant, any change in accommodation when a subject successfully fused the stimulus must be driven primarily by vergence. With a typical 63 mm interpupillary

distance (Dodgson 2004; Gordon et al. 1989), the target was presented at a vergence/accommodation distance of 67 cm (i.e. 1.50D).

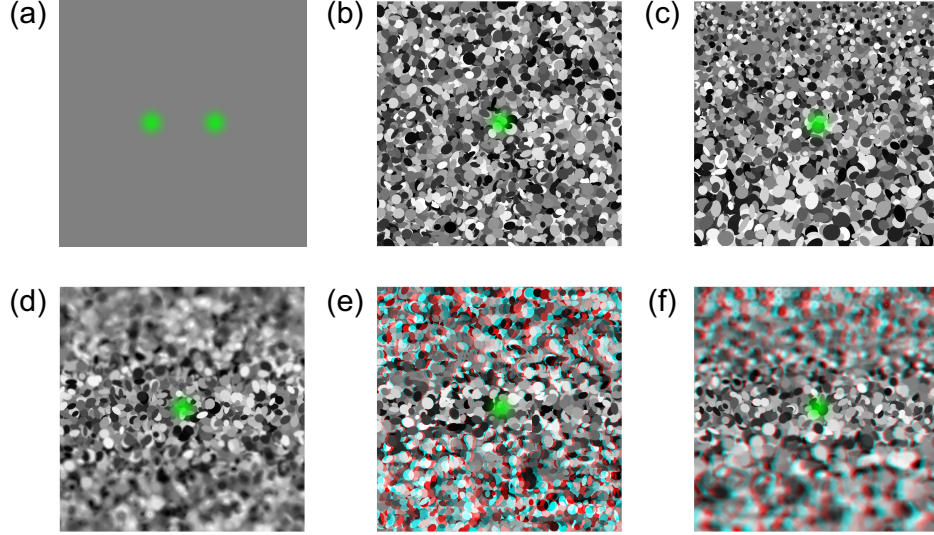


Figure 7.2: **Stimuli to drive vergence and accommodation.** (a) In Step 1, a Gaussian blurred dot with -3.5 degrees of uncrossed disparity was used to drive vergence behind the surface of the display. Its blurred structure eliminated spatial cues for accommodation. (b-f) In Step 2, dead leaves stimuli were used to drive vergence and accommodation to the surface of the display. At central fixation, zero disparity and sharply focused images specified vergence and accommodation at the depth of the monitor, and the potential influence of the visual periphery was assessed by presenting 5 simulated depth cue configurations in the peripheral visual field: (b) No Depth Cues (c) Relative Size gradient (d) Blur gradient (e) Disparity gradient (f) Size, Blur and Disparity gradient. In figures (e,f) stereoscopic disparity may be viewed using red-cyan anaglyph glasses. See main text for a detailed description of stimuli and stimulus generation procedures.

## **Step 2: Vergence and Accommodation Responses with Peripheral**

**Depth Cues** Vergence and accommodation to the surface of the display were driven by 40x40 degree patches of “dead leaves” (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001). Example stimuli are shown in Figures 2b-f. At the center of the stimulus, the same 0.5 degree green fixation dot as used in Step 1 was shown with zero disparity to drive vergence to the surface of the display. The dead leaves stimuli were constructed from a set of 3000 ellipses each assigned a center position, orientation, aspect ratio, and luminance drawn from pseudo-random uniform distributions. The mean luminance of the ellipses was equal to the display background (25 cd/m<sup>2</sup> through the shutter glasses) and the rms contrast was 0.5.

The size, blur and disparity of each ellipse was independently controlled to provide one or more simulated depth cues to simulate a surface slanted in depth with the superior area of the surface tilted away from the observer and the inferior area titled towards the observer:

**Zero gradient (Figure 7.2b)** : mean ellipse size was constant (0.5 degrees - 1.5 degrees), ellipse edges were sharp and disparity was zero. This stimulus appeared as a flat surface on a fronto-parallel plane anterior to the observer.

**Size gradient (Figure 7.2c)** : ellipse edges were sharp and disparity was zero. Mean ellipse size was constant (0.5 degrees - 1.5 degrees) at the horizontal center of the stimulus and linearly decreased in both the top and the bottom of the stimulus.

**Blur gradient (Figure 7.2d)** : mean ellipse size was constant (0.5 degrees

- 1.5 degrees) and disparity was zero. Ellipse edge blur increased linearly above and below the fixation height. Individual ellipses were filtered in the frequency domain using a sinc filter  $\Gamma(\omega)$  defined as:

$$\Gamma(\omega) = \frac{\sin(\frac{\pi\omega}{\lambda})}{\frac{\pi\omega}{\lambda}} \quad (7.1)$$

Where  $\lambda$  determines the spatial frequency,  $\omega$ , of the first phase reversal. To obtain a linear increase in blur away from the fixation height, the inverse of  $\lambda$  was linearly increased from 0.4 to 8 degrees/cycle. The Sync filter was chosen because it both attenuates high spatial frequencies and introduces phase reversals typical of the modulation transfer function of an optical system with a circular aperture such as the human pupil (Murray and Bex 2010). Because of the tilt shift effect (Vishwanath and Blaser 2010), a stimulus with increasing levels of blur above and below fixation should be perceived as a near surface slanted in depth, with the top tilted away from the observer.

**Disparity gradient (Figure 7.2e) :** mean ellipse size was constant (0.5 degrees - 1.5 degrees), and ellipse edges were sharp. Ellipse disparity increased from -1.4 degrees at the bottom (crossed disparity, ellipses before the screen) to +1.4 degrees at the top (uncrossed disparity, ellipses behind the screen).

**Size, Blur and Disparity gradients (Figure 7.2f) :** ellipse size, blur and disparity varied together as in conditions b), c) and d).

Note that only conditions a) and f) contained no depth cue conflicts; the mean simulated depth of all depth gradients was centered on the display

surface; and the horizontal mid-point of all stimuli was the same in all cases and specified depth on the surface of the display.

## Procedure

Subjects who met inclusion criteria were asked to wear stereoscopic shutter-glasses and place their heads into the chinrest in front of the 3D monitor. Subjects underwent an initial training session of 5 trials in which all stimuli and tasks were demonstrated.

On each trial, subjects first completed Step 1, Vergence driven Accommodation to Simulated Disparity, then completed Step 2, Vergence and Accommodation Responses with Peripheral Depth Cues. Subjects fused the blurred green fixation dot rendered with -3.5 degrees of uncrossed disparity on a uniformly gray background. When satisfied they were fusing the stimulus, subjects pressed a key on the keyboard in front of them, which initiated Step 1. The uncrossed disparity fixation dot was kept on screen for 0.5 seconds, and then Step 2 began where the disparity of the fixation dot was set to zero and one of the 5 dead leaves stimuli (Figure 7.2) was presented, at random across trials. The fixation dot and the horizontal center of the dead leaves stimuli drove vergence and accommodation to the surface of the display. Observers were instructed to maintain steady fixation onto the fixation dot at the center of the screen. After 5 seconds, the dead leaves stimulus was replaced once again by the disparity defined fixation dot onto a uniformly gray background, which signaled the end of the trial.

Once the subject had completed the training session, the experiment proceeded to the main test session. The test session was identical to the training

session except that the test session consisted of 10 repetitions of each of the 5 peripheral cue conditions in random order, thus each subject completed 50 trials.

## Data Analyses

For each subject, raw data were rescaled. Mean accommodation and mean vergence were defined as baseline zero at the monitor distance when the subjects were viewing the Zero gradient dead leaves stimulus (Step 2, Zero gradient condition). Thus, all data are relative to the baseline level of accommodation when observers are viewing the flat image on the computer screen, which is like a typical computer use situation. Therefore, all results reported in this chapter are relative to the vergence/accommodative state of each observer when viewing the 2D monitor, and no conclusions may be drawn as to the absolute values of vergence and accommodation.

Furthermore, negative vergence values were defined here to signify vergence to uncrossed disparities (i.e. vergence to simulated depth beyond the surface of the monitor). Negative accommodation values instead signify accommodation behind the surface of the monitor (because diopters =  $1/\text{meters}$ ).

Both vergence and accommodation data were separately fit with a piecewise function, composed of a constant and an exponential (Beers and Van Der Heijde 1994; Beers and Van der Heijde 1996; Kasthurirangan and Glasser 2005b; Kasthurirangan, Vilupuru, and Glasser 2003; Yamada and Ukai 1997), defined as:

$$D(t) = \begin{cases} d_0 & : t < t_0 \\ d_0 + a(1 - e^{-(t-t_0)/\tau}) & : t \geq t_0 \end{cases} \quad (7.2)$$

where  $d_0$  was the level of vergence/accommodation during Step 1 of each trial, when observers were fixating the blurred fixation dot with -3.5 degrees of uncrossed disparity on a blank background.

For each subject I also separately computed the variability of accommodation and vergence during Step 1 as the standard deviation of the data prior to time  $t=0$ .

Time  $t=0$  was defined as the start of Step 2; i.e. the onset of the dead leaves stimuli, during which the disparity of both the fixation and the center of the dead leaves stimuli was zero. The decreasing exponential function characterized the response with a delay  $t_0$ , a response amplitude  $a$ , and an exponential half-life  $\tau$ . An example of a function fitted to data from one condition from a representative observer can be seen in Figure 7.3. For each subject and each condition, for vergence and accommodation separately, I also computed the steady state response variability, which was defined as the standard deviation of the data from the time point  $t_{ss} = t_0 + \tau \times \log 20$ , at which the exponential had reached 95% steady state. For each subject, accommodation data from the left and right eye were fit to a single function.

The refractive status data regarding level ( $d_0$ ) and variability of vergence and accommodation during Step 1 were analyzed with a one-way between subjects ANOVA. All data from Step 2, meaning parameters  $a$ ,  $t_0$ , and  $\tau$  of the exponential fits as well as steady state vergence and accommodation variability, were analyzed with 2 (refractive status, between subjects factor) x 5 (cue condition, within subjects factor) mixed design ANOVA.



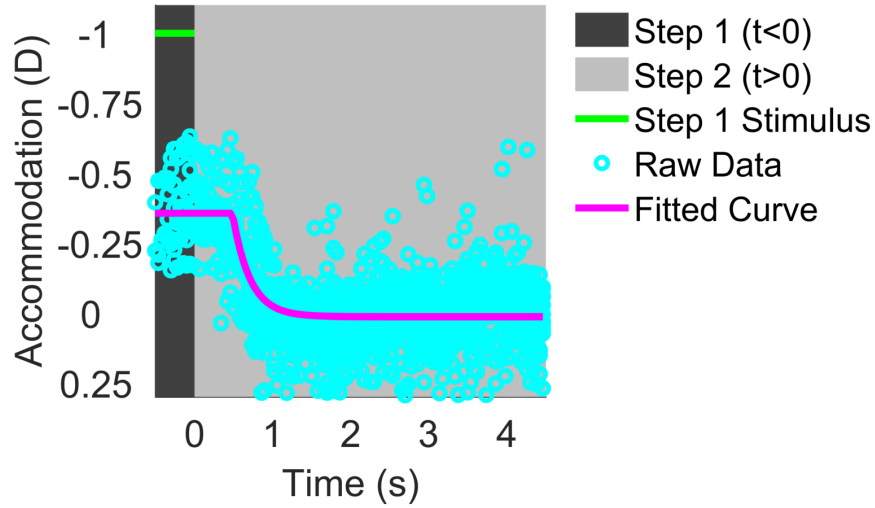


Figure 7.3: **Accommodation as a function of time for a representative observer in which Step 2 was a “Zero gradient” stimulus.** Cyan circles represent raw accommodation data and the magenta line is the best fitting accommodation curve (equation 2). Step 1 (Vergence driven Accommodation to Simulated Disparity, dark gray shaded area) occurs before time  $t=0$ s and specifies a target at 1 diopter away from the monitor (green line). Step 2 (Vergence and Accommodation Responses with Peripheral Depth Cues, light gray shaded region) begins at time  $t=0$ s, and specifies a target on the surface of the monitor.

## 7.3 Results

Figure 7.4 presents group vergence responses as a function of time for myopes (red) and emmetropes (blue). Negative time points in each panel (dark gray shaded region) correspond to data for Step 1. Time 0 indicates the start of Step 2 (light gray shaded region) for each of the 5 depth gradient conditions (a-e). Red and blue shaded areas represent the 95% confidence intervals of the mean vergence trace, and solid lines show the mean best fitting piecewise function.

Figure 7.5 presents group accommodation responses as a function of time for myopes (red) and emmetropes (blue). Negative time points (dark gray shaded region) in each panel correspond to data for Step 1. Time  $t=0$  indicates the start of Step 2 (light gray shaded region), for each of the 5 depth gradient conditions (a-e). Blue and red shaded areas are 95% confidence intervals of the mean accommodation trace, and full lines show the average fitted piecewise function.

### **Step 1: Vergence driven Accommodation to Simulated Disparity**

Negative time points in Figure 7.4 show that both emmetropes and myopes verged away from the monitor depth in response to the disparity of the fixation target. Vergence during Step 1 was not significantly different between myopes and emmetropes ( $F_{1,19}=3.0$ ,  $p=0.10$ ). The variability of vergence during Step 1 was also not significantly different between myopes and emmetropes ( $F_{1,19}=2.4$ ,  $p=0.14$ ).

Negative time points in Figure 7.5 indicate that both groups, emmetropes and myopes, accommodated away from the monitor depth in response to the

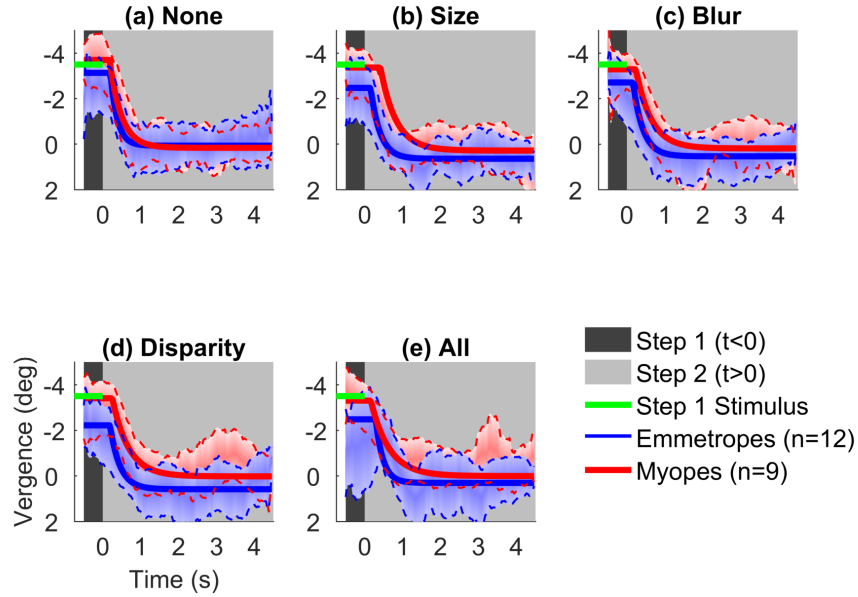


Figure 7.4: **Vergence response curves for Steps 1 and 2.** Shaded regions bounded by dotted lines are 95% bootstrapped confidence intervals of mean vergence as a function of time for myopes (red) and emmetropes (blue). Negative time points correspond to Step 1, and time 0 indicates the start of Step 2. Filled lines are piecewise functions fitted to the raw data and then averaged across observers in each refractive group and each peripheral cue condition: (a) Zero gradient (b) Size gradient (c) Blur gradient (d) Disparity gradient (e) Size, Blur and Disparity gradients. The convention employed in this chapter places the monitor at 0 degrees of vergence, with divergence beyond the surface of the monitor specified as increasingly negative values. The disparity defined fixation dot in Step 1 specifies vergence at -3.5 degrees away from the surface of the monitor (here highlighted by the green lines).

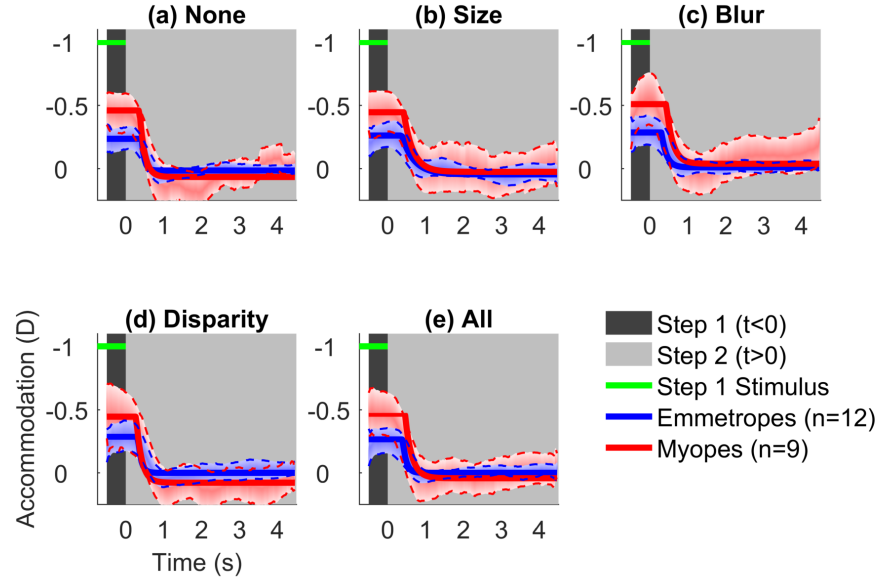


Figure 7.5: **Accommodation response curves for Steps 1 and 2.** As in Figure 7.4 except for Accommodation Responses. Shaded regions bounded by dotted lines are 95% bootstrapped confidence intervals of mean accommodation for myopes (red) and emmetropes (blue). Filled lines are fitted piecewise functions averaged across observers in each refractive group and each peripheral cue condition: (a) Zero gradient (b) Size gradient (c) Blur gradient (d) Disparity gradient (e) Size, Blur and Disparity gradients. Negative time points refer to Step 1, where accommodation is specified by vergence stimulus to be 1D behind the display. Step 2 began at  $t=0$ s and the fixation and center of the stimulus specified accommodation at the distance of the monitor.

change in the disparity of the fixation target. Accommodation changes were consistent with an increase in depth of the fixation target in the absence of a reliable blur cue, in line with the vergence-accommodation linkage. However, although the fixation target was 1.00D away from the monitor, accommodation change was less than 1.00D in both refractive groups.

Figure 7.6a shows accommodation away from the surface of the display during Step 1 for myopes (red) and emmetropes (blue). Myopes on average accommodated 0.47D [-0.31, -0.61; 95% CI] away from the display surface, whereas emmetropes accommodated -0.27D [-0.20, -0.37; 95% CI], a 43% smaller accommodative response ( $F_{1,19}=5.13$ ,  $p=0.036$ ). Given that the cue to depth change was disparity-driven vergence, the accommodation response was primarily driven by the vergence response. The larger accommodative response of myopes suggests that the vergence-accommodation link is stronger in myopes than in emmetropes.

Figure 7.6b shows the variability of accommodation during Step 1. Mean accommodation variability was 0.17D [0.15, 0.22; 95% CI] for emmetropes and 0.24D [0.22, 0.26; 95% CI] for myopes. Thus, myopes exhibited a 41% increase in accommodation variability with respect to emmetropes ( $F_{1,19}=12.91$ ,  $p=0.0019$ ).

**Step 2: Vergence and Accommodation Responses with Peripheral Depth Cues** There was no significant difference in the amplitude of the vergence response of observers across refractive groups ( $F_{1,19}=1.62$ ,  $p=0.22$ ) or gradient depth cue conditions ( $F_{4,76}=1.13$ ,  $p=0.35$ ), and there was no significant interaction between refractive group and cue condition ( $F_{4,76}=0.19$ ,

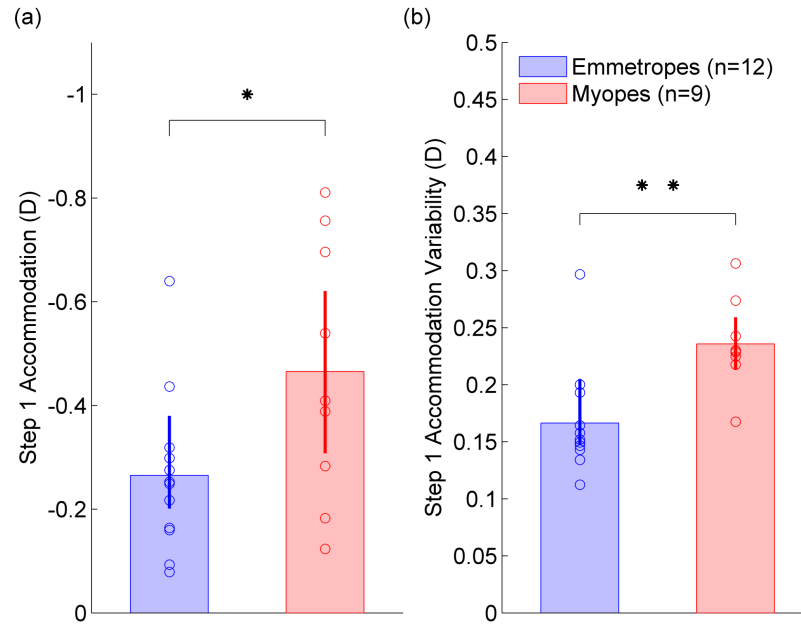


Figure 7.6: **Accommodation and accommodation variability during Step 1.** (a) Mean value of  $d_0$  of the fitted accommodation response curves averaged across emmetropes (blue) and myopes (red). (b) Mean variability of accommodation during Step 1 averaged across emmetropes (blue) and myopes (red). Error bars are 95% bootstrapped confidence intervals. Circles are individual subject data. \*  $p<0.05$ , \*\*  $p<0.01$

$p=0.94$ ). Vergence responses were initiated on average 0.22 seconds [0.18, 0.27; 95% CI] after the dead leaves stimulus onset. This response latency did not vary across the refractive groups ( $F_{1,19}=1.0$ ,  $p=0.33$ ) or cue condition ( $F_{4,76}=0.66$ ,  $p=0.62$ ) and the interaction between refractive groups and cue condition did not quite reach significance ( $F_{4,76}=2.4$ ,  $p=0.057$ ). Thus, the presence of peripheral depth cues had little or no effect on the onset or endpoint of vergence responses in myopes or emmetropes.

The exponential half-life of the fitted vergence response curves indicates the speed of the vergence response, with smaller values indicating faster responses. Mean,  $\tau$  was 0.23 seconds [0.17, 0.31; 95% CI] for emmetropes and 0.41 seconds [0.27, 0.57; 95% CI] for myopes. ANOVA analysis on the exponential half-life parameter revealed that myopes had a significantly slower responses (78% larger  $\tau$ ) than emmetropes ( $F_{1,19}=4.55$ ,  $p=0.046$ ), as shown in Figure 7.7. ANOVA analyses on the exponential half-life parameter found no main effect of cue condition ( $F_{4,76}=0.48$ ,  $p=0.75$ ), and no significant interaction between refractive group and cue condition ( $F_{4,76}=0.27$ ,  $p=0.90$ ).

Finally, vergence steady state variability was the same across refractive groups ( $F_{1,19}=1.74$ ,  $p=0.20$ ) and cue condition ( $F_{4,76}=0.79$ ,  $p=0.54$ ), with no significant interaction between refractive group and cue condition ( $F_{4,76}=0.95$ ,  $p=0.44$ ).

Because myopes relaxed accommodation to a greater extent than emmetropes during Step 1, myopes also had to execute larger accommodation responses to changes in target disparity ( $F_{1,19}=8.63$ ,  $p=0.0085$ ). However, there was no overall effect on accommodation amplitude change of the specific peripheral depth cue condition ( $F_{4,76}=0.2$ ,  $p=0.94$ ), and no signifi-

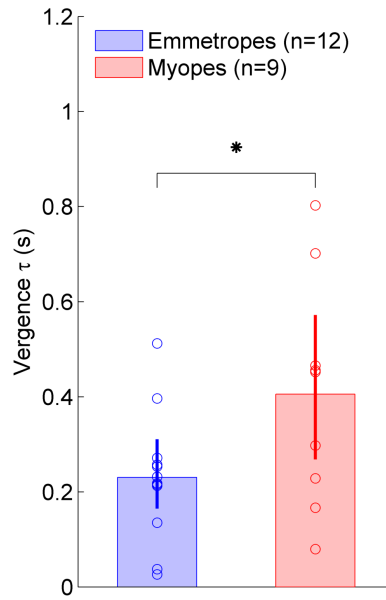


Figure 7.7: **Exponential half-life  $\tau$** . Exponential half-life  $\tau$  of the fitted vergence response curves averaged across myopes (red) and emmetropes (blue) and across cue conditions. Error bars are 95% bootstrapped confidence intervals. Circles are individual subject data. \*  $p < 0.05$



cant interaction between refractive group and cue condition ( $F_{4,76}=0.80$ ,  $p=0.53$ ). The accommodation responses were initiated on average 0.37 seconds [0.29-0.43, 95% CI] after the dead leaves stimulus onset. This response latency did not vary across refractive group ( $F_{1,19}=0.14$ ,  $p=0.71$ ) or cue condition ( $F_{4,76}=1.29$ ,  $p=0.28$ ) and no significant interaction was observed ( $F_{4,76}=0.82$ ,  $p=0.52$ ). The exponential half-life parameter of the accommodation response was on average 0.13 seconds [0.08-0.20, 95% CI]. ANOVA analysis of the exponential half-life parameter of the accommodation response showed no main effect of refractive group ( $F_{1,19}=0.33$ ,  $p=0.57$ ), no main effect of cue condition ( $F_{4,76}=1.07$ ,  $p=0.38$ ), and no significant interaction between refractive group and cue condition ( $F_{4,76}=1.38$ ,  $p=0.25$ ). Thus, the presence of peripheral depth cues did not affect the accommodative responses of myopes or emmetropes.

Lastly, Figure 7.8 shows average accommodation steady state variability for myopes and emmetropes in each cue condition. ANOVA analysis found a significant main effect of refractive group ( $F_{1,19}=6.24$ ,  $p=0.022$ ), a significant main effect of cue condition ( $F_{4,76}=5.1$ ,  $p=0.0011$ ), and a significant interaction between refractive group and cue condition ( $F_{4,76}=2.7$ ,  $p=0.037$ ). I thus compared myopes to emmetropes for each cue condition via Wilcoxon rank sum tests. Myopes had significantly higher steady state variability than emmetropes for the no cue condition (53% increase,  $Z=2.38$ ,  $p=0.017$ ), relative size cue condition (43% increase,  $Z=2.24$ ,  $p=0.025$ ), blur cue condition (37% increase,  $Z=2.38$ ,  $p=0.017$ ), and the disparity cue condition (48% increase,  $Z=2.38$ ,  $p=0.017$ ). In the all cue condition the difference in accommodation steady state variability between myopes and emmetropes

was not quite statistically significant (25% increase,  $Z=1.81$ ,  $p=0.070$ ). Note that the presence of one outlier in the myopic group with much higher than average accommodation variability did not affect the pattern nor the statistical significance of the results.

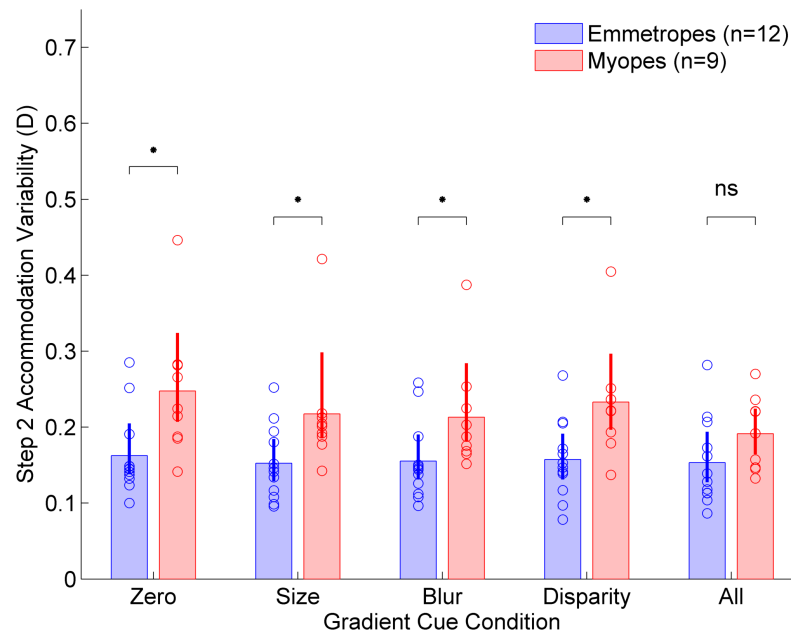


Figure 7.8: **Accommodation steady state variability.** Accommodation Steady State Variability averaged across myopes (red) and emmetropes (blue) for each cue condition. Error bars are 95% bootstrapped confidence intervals. Circles are individual subject data. \*  $p<0.05$

## 7.4 Discussion

When viewing blurred targets presented with binocular disparity on a stereoscopic 3D monitor, myopes and emmetropes executed vergence eye movements of comparable magnitude to the stimuli. However, while the latency of vergence eye movements was not significantly different between refractive groups, the time constant of the vergence response in myopic observers was significantly slower than in emmetropes. These observations are consistent with Vienne et al. (2014) who found that vergence responses were slower with conflicting disparity and blur cues, as well as with Semmlow and Wetzel (1979) who showed that binocular vergence movements are faster when disparity and blur are both available than when only disparity specifies a change in distance. Thus, slower vergence movements in myopes are consistent with the previous finding that myopes make less use of retinal defocus information (Cufflin, Mankowska, and Mallen 2007; Rosenfield and Abraham-Cohen 1999) and extends this notion from the perceptual to the oculomotor system.

I employed blurred target stimuli located behind the monitor in order to degrade defocus blur cues to image depth and thus any change in accommodation was primarily driven by a change in vergence. Accommodation responses were consistent with the change in disparity of the stimuli, providing clear evidence for the well-known coupling between vergence and accommodation (Fincham and Walton 1957). While vergence eye movements to stimuli containing binocular disparity were generally comparable between emmetropes and myopes, there were significant differences in accommodative responses to these stimuli. The magnitude of accommodative change with vergence change was significantly larger in myopes than in emmetropes. This suggests

that the vergence-accommodation coupling is stronger in myopes than in emmetropes under the present conditions of degraded defocus information. It is possible that this is associated with evidence that myopes are less sensitive to image defocus than emmetropes (Cufflin, Mankowska, and Mallen 2007; Rosenfield and Abraham-Cohen 1999).

I fit the data collected for this chapter with a commonly employed exponential function to facilitate comparisons between these results and the existing literature. Indeed, the values of the fitted parameters in this study nicely conform to the literature describing the characteristics of the accommodative response (H. A. Anderson et al. 2010; Beers and Van Der Heijde 1994; Beers and Van der Heijde 1996; Kasthurirangan and Glasser 2005b; Kasthurirangan, Vilupuru, and Glasser 2003; Yamada and Ukai 1997). However, it is worth noting that the onset of both vergence and accommodative response is unlikely to be as abrupt as the exponential model imposes. Thus, the latency of the vergence and accommodative responses is likely to be slightly overestimated when employing this exponential model.

Adult myopes have been shown to have more variable accommodation with high contrast targets viewed monocularly (Day et al. 2006; Seidel, Gray, and Heron 2003) but not binocularly (Seidel, Gray, and Heron 2005). Adult myopes also had more variable accommodation when reading binocularly (Harb, Thorn, and Troilo 2006). In agreement with the more naturalistic reading task (Harb, Thorn, and Troilo 2006) but not with cross targets fixated binocularly (Seidel, Gray, and Heron 2005), I found that myopes have less stable accommodation than emmetropes when binocularly viewing high contrast, naturalistic depth defined stimuli. More specifically, when depth

cues specified a flat perpendicular surface, and when depth cues were conflicting (some specifying a flat perpendicular surface and some specifying a surface tilted in depth), myopes exhibited higher accommodation variability than emmetropes. When instead a slanted surface in depth was simulated coherently by all cues, myopes and emmetropes exhibited statistically similar levels of accommodation variability, although on average myopes still exhibited more variable accommodation.

Previous work has shown that progressing myopic children underaccommodate to defocus but not to a target that changes its distance (Abbott, Schmid, and Strang 1998; Gwiazda et al. 1993; Jiang 1997), and that accommodation in myopes returns to the level of emmetropes once myopia progression has stopped (Gwiazda et al. 1995). The improvement in accommodation may involve an adaptation process within the accommodative system that upregulates the gain of accommodation (Gwiazda et al. 1995; Vera-Diaz et al. 2004). An increase in the gain of the myopic accommodative system is consistent with the increase the variability of accommodation observed in this study.

These findings are all consistent with the notion that myopes are less sensitive to defocus cues and lend strength to the hypothesis that greater retinal blur signals may be an important factor in the development of myopia. Specifically, both vergence and accommodation differed across refractive groups when viewing simulated depth content on stereoscopic displays. Myopes executed slower vergence eye movements, exhibited a stronger vergence-accommodation link, and had more variable accommodation overall.

Investigating these differences is necessary to understand whether the advance of virtual reality technology might further exacerbate the increasing incidence of myopia, and whether we might turn this technology to our advantage to instead oppose the rise of refractive error development. One potential avenue of investigation in this regard is that inaccuracies of accommodation when viewing flat images (typical of indoor activities such as reading or computer use) may cause long-term blur on the retina, which may impede precise emmetropization. Having rich depth structure presented on virtual reality displays may instead be beneficial to stabilizing accommodation and promote normal ocular growth. To this end, virtual reality displays with coherent or near-correct depth and focus cues may be implemented in gaze contingent fashion (Duchowski et al. 2014; Maiello et al. 2014; Mauderer et al. 2014) with newly available low-cost eye tracking systems (Dalmaijer 2014; Gibaldi et al. 2016), with switchable lens systems (Love et al. 2009), and even as head mounted displays (see Chapter 12, Chessa, Maiello, Borsari, et al. 2016 and Liu and Hua 2010).

Future work should thus initially focus on understanding defocus sensitivity in virtual reality displays. The method I have devised employs simulated disparity to drive and continuously monitor accommodation. This method may be employed in future research to study natural dioptric defocus perception using an observer’s own optics without the need to construct custom volumetric (MacKenzie, Hoffman, and Watt 2010) or multiplane display rigs (Sebastian, Burge, and Geisler 2015). Future investigations should also include children to evaluate possible causative or preventative effects of simulated visual environments on refractive error development.

## Part 3: Perception, Action and Motion

## Chapter 8

# Monocular and Binocular Contributions to Oculomotor Plasticity

### 8.1 Introduction

Humans and many animals with overlapping binocular visual fields continuously shift their gaze point in three-dimensional (3D) space in order to bring targets of interest onto the high-resolution foveae of the two eyes. These gaze shifts require highly precise and coordinated fast saccadic and slow vergence eye movements. To achieve binocular coordination in a 3D environment, Hering (1868) proposed that both eyes are innervated by conjugate command signals. Alternatively, Von Helmholtz (1867) argued that binocularly coordinated visual systems evolved from independently moving lateral eyes and that each eye is controlled independently (Zhou and King 1998) (for a



review see King 2011). The finding that some saccadic eye movements in 3D space result in each eye moving by a different amount (Enright 1992; Maxwell and King 1992) has generally been seen as a challenge to Hering’s law of equal innervation. However, it has been argued these data may yet be explained by a vergence command superimposed on a conjugate saccadic command (Ramat et al. 1999).

In order to maintain the accuracy and precision of eye movements in 3D environments over time, oculomotor control processes must also be capable of adapting to lifespan changes that affect the forces acting on each eye. For example during normal aging, extraocular muscles become displaced (Clark and Demer 2002), loss of structural support leads to orbital fat prolapse (Salvi, Akhtar, and Currie 2006), the area of the orbital aperture increases and the eye sockets become wider and longer (Shaw et al. 2011). The plasticity of the oculomotor control system that is required to correct for such changes has been widely investigated with short-term saccade adaptation paradigms (McLaughlin 1967). In a double-step saccade adaptation study, an observer is asked to make a saccadic eye movement to a target. During the saccade, the target is displaced so that the observer’s saccade fails to bring their fovea to the target. Due to saccade masking (Bridgeman, Hendry, and Stark 1975; Deubel, Wolf, and Hauske 1985; Deubel, Schneider, and Bridgeman 1996; Deubel, Bridgeman, and Schneider 1998; Deubel, Schneider, and Bridgeman 2002) and poor sensitivity to high speed retinal images (Dorr and Bex 2013; Volkman 1962), the intra-saccadic displacement is not noticed by the observer. Consequently, when the eyes land, there is an apparent sensorimotor mismatch between the actual foveal signal and the foveal signal expected

given the executed eye movement (Collins and Wallman 2012). Over the course of a few dozen trials, the oculomotor system recalibrates, altering the saccade gain so as to reduce the mismatch on subsequent trials (for a recent review see Pélisson et al. 2010). Analogous saccadic adaptation processes could in principle account for longitudinal oculomotor plasticity during aging. Additionally, in order to maintain accurate binocular calibration when aging processes around each eye occur asymmetrically, saccade adaptation would have to operate independently for each eye. However, it is not currently known whether oculomotor adaptation can occur separately for each eye, since adaptation is usually studied with identical error signals in both eyes. One study that investigated monocular adaptation showed that adaptation induced only in one eye transfers to the other eye (Albano and Marrero 1995), which would suggest that saccade adaptation may not be capable of operating independently for each eye.

In this chapter, I employed the classic short-term double step saccade adaptation paradigm (McLaughlin 1967) to induce saccade recalibration in six normally-sighted young adults. I modified this paradigm to test how the oculomotor system adapts to different errors in the saccade end points of the two eyes. I compared oculomotor responses to conjugate saccade adaptation (i.e. same error in both eyes) with oculomotor responses to dichoptic saccade adaptation (i.e. different error in the two eyes). By studying asymmetric monocular oculomotor adaptation, I addressed two outstanding questions in visual neuroscience. Firstly, I examine whether eye movement control is conjugate for both eyes, as originally argued by Hering (1868), or independent for each eye, as originally suggested by Von Helmholtz (1867). If we find

asymmetric adaptation, this would provide convincing evidence in favour of independent monocular innervation. Secondly, I examine whether independent saccade adaptation could provide the basis for oculomotor calibration for asymmetric changes on the forces of the eyes across the lifespan.

The experimental conditions are illustrated in Figure 8.1. Observers executed a saccade to track a Gabor target ( $\sigma = 0.5$  degrees,  $\omega = 4$  cycles/degree) that stepped from the centre of the monitor to a position at 8 degrees to the left (Figure 8.1a) or right. During the saccade, the target stepped invisibly a second time by 10% (0.8 degrees) of the initial displacement. As shown in Figure 8.1b, the secondary target step was either outward in both eyes (outwards step), inward in both eyes (inwards step), or outward in the temporally moving eye and inward in the nasally moving eye (dichoptic step). For the outward and inward step conditions, observers' saccade gain should increase and decrease, respectively (Pélisson et al. 2010). It is currently unknown how the oculomotor system will respond to the dichoptic steps.

## 8.2 Methods

### Participants

Two experienced and five naïve observers (all male, ages ranging from 20 to 32 years) were recruited for the study. All subjects had normal or corrected to normal vision. Six out of the seven subjects had had normal stereo vision, as confirmed by a stereoscopic Vernier acuity task performed prior to the main experiment. One subject had mild amblyopia, which was detected by

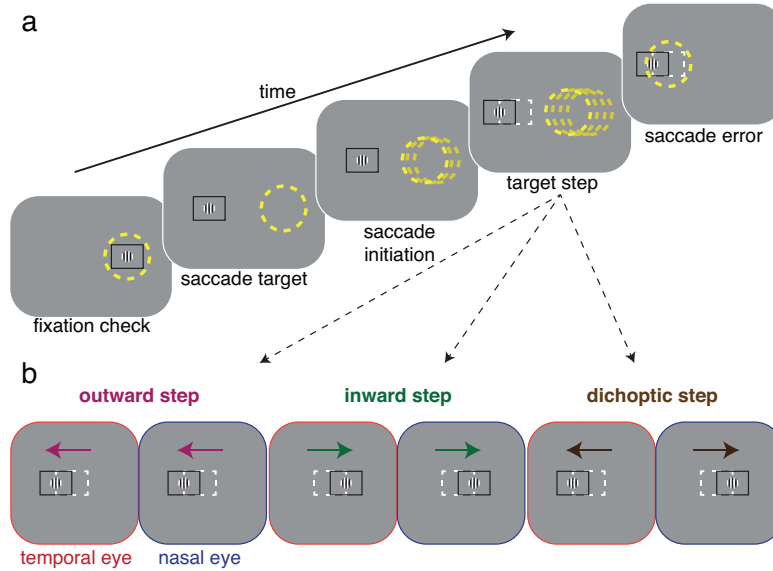


Figure 8.1: **Saccade adaptation paradigm.** (a) Schematic of a trial. The yellow dashed circle represents an observer’s hypothetical gaze location. The observer initially fixates a target Gabor, framed by a black nonius square to aid binocular fusion (Shimono et al. 1998). Once fixation is verified by the eye tracker, the target is shifted 8 degrees leftward (or rightward across blocks), prompting the observer to make a leftward saccade. Saccade initiation is detected online (see Data Analysis) and, during the saccade, the intra-saccadic target step displaces the target from its pre-saccadic position (dashed white box, not present in the stimulus). In this example, the saccade lands at the pre-saccadic target position, resulting in an oculomotor error signal. (b) Three experimental conditions. The intra-saccadic target step could be outward in both eyes (outward step), inward in both eyes (inward step), or outward in the temporally moving eye and inward in the nasally moving eye (dichoptic step).

the stereoscopic acuity task, and was excluded from further testing.

## Stimuli and Apparatus

All experiments were programmed with the Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) and Eyelink Toolbox (Cornelissen, Peters, and Palmer 2002) in Matlab (MathWorks). Eye-tracking was performed using the EyeLink 1000 (SR Research) desktop mount eye-tracker. The eye tracker was calibrated binocularly using the native thirteen-point calibration routine at the start of each experiment. Eye-tracking data was recorded at 500 Hz per eye and stimuli were presented on an ASUS VG278HE LCD monitor with a resolution of 1920x1050 pixels (display dot pitch 0.311 mm) at 120 Hz. The system latency was 24msec, measured with a video-based method (Saunders and Woods 2014), which is below the duration of a typical saccadic eye movement of the amplitude measured in the present study. The monitor was run from an NVidia Quadro FX 4600 graphics processing unit. Observers were seated 50cm in front of the monitor with their heads stabilized in a chinrest and wore active stereoscopic shutter-glasses (NVIDIA 3D Vision) during all experiments to control dichoptic stimulus presentation. The cross talk of the dichoptic system was 1% measured with a Spectrascan 6500 photometer.

## General Procedure

Schematics of the experimental design are shown in Figure 8.1. An observer's task was to make saccades directly to a Gabor target ( $\sigma = 0.5$  degrees,  $\omega = 4$  cycles/degree, 55% contrast) immediately after it stepped 8 degrees to the

left or to the right of initial central fixation. Each observer completed two sessions for each adaptation condition (outwards, inwards and dichoptic). In one session saccades were always to the left, in the other session saccades were always to the right. The order of conditions and sessions was counterbalanced across observers via a Latin square design. Within a session, an observer completed 175 trials: the first 25 trials were Baseline Trials, in which there was no intrasaccadic target displacement; trials 26 to 100 were Adaptation Trials, in which a different intrasaccadic shift was introduced depending on the condition; and trials 101 to 175 were Recovery Trials, in which the intrasaccadic target displacement was no longer presented.

At the start of each trial, subjects fixated the central Gabor target encompassed by a 4x3 (width x height) degree nonius bounding box. The nonius box served to aid vergence at the stimulus depth (Shimono et al. 1998). Subjects were required to make a correct stereoscopic vernier judgment in order for the trial to commence. This was performed to ensure that subjects were correctly fixating and attentive. Following the stereoscopic acuity judgment, subjects were required to maintain fixation within the central 1.5 degrees (with compliance monitored by the eye-tracker). After a random latency drawn from a uniformly distributed time interval (1000-1500 ms), the central Gabor was abruptly shifted 8 degrees to either the left or right, depending on the session. Subjects were instructed to immediately saccade to the new eccentric target location and to maintain steady fixation on the eccentric target. After 1000 ms the target disappeared and the central red dot reappeared, indicating that a new trial could be commenced.

During the 25 Baseline Trials and the 75 Recovery Trials, the target

shifted by 8 degrees in both eyes with no further change in position. During the 75 Adaptation Trials, following the first 8 degree displacement, the target was displaced a second time by 0.8 degrees. The direction of the secondary shift was outwards in both eyes (outwards condition), inwards in both eyes (inward condition), or outwards in the temporally moving eye and inwards in the nasally moving eye (dichoptic condition). This secondary target displacement took place as soon as the eye-tracker detected that the mean position of the two eyes had moved 2 degrees from central fixation. The secondary displacement thus occurred during the saccade, ensuring that the secondary displacement was not detected (Bridgeman, Hendry, and Stark 1975; Deubel, Wolf, and Hauske 1985; Deubel, Schneider, and Bridgeman 1996; Deubel, Bridgeman, and Schneider 1998; Deubel, Schneider, and Bridgeman 2002; Dorr and Bex 2013; Volkman 1962). After completing all experimental sessions (individual sessions were spread out over  $\sim 2$  weeks of testing, each observer completed a total of 1050 trials), naïve observers were debriefed and in all cases were surprised to learn about the experimental manipulation: none had noticed the secondary displacement. During the dichoptic adaptation trials, all observers were aware that the end stimulus contained depth information, since the manipulation introduced a small amount of uncrossed disparity, even though they were unaware of the intrasaccadic displacement.

## Data Analysis

Saccades were detected via the Eyelink software in its default configuration, which classifies data samples as belonging to saccades when these exceed a velocity threshold of 30 degrees/second, an acceleration threshold of 8000

degrees/second<sup>2</sup>, and a motion threshold of 0.1 degrees. Raw saccade amplitude data were smoothed with a lowess regression with a span of 15 trials for illustrative purposes. The data were fit with a second-degree polynomial (parabolic) equation:

$$S_{amp} = R \times n^2 + D \times n + I \quad (8.1)$$

where  $S_{amp}$  is the saccade amplitude,  $n$  is the trial number,  $R$  is the rate of change of the parabola, and  $D$  is the declivity of the parabola at the y-axis intercept  $I$ . Saccade amplitude data and fitted parameter estimates were averaged across sessions and observers. Parameter estimates were analysed using a 2 (eye) x 3 (adaptation condition) within-subject ANOVA and, where appropriate, post-hoc analyses were performed using a Tukey-Kramer single-step, multiple comparison procedure.

Vergence was computed with reference to the display screen as the difference in left and right horizontal eye coordinates (i.e., vergence was zero when both eyes were pointing at the same spot on the monitor surface). The transient divergence typically observed during saccades (Collewijn, Erkelens, and Steinman 1997) was removed from the vergence traces measured during the adaptation trials. This was accomplished by subtracting from the vergence traces measured during the adaptation trials the average vergence trajectory observed during the Baseline trials. Vergence data were then fitted with an asymmetric logistic function with equation:

$$V(t) = L + \frac{H - L}{(1 + e^{-R \times (t - t_0)})^{10}} \quad (8.2)$$

Where  $t$  is time from saccade onset and  $V(t)$  is the vergence response.



The amplitude of the vergence response was defined as  $V_{amp} = H - L$ . The time from saccade onset at which a vergence eye movement was commenced was estimated to be the time at which the vergence response had increased from baseline by 1/100th of  $V_{amp}$  :

$$V(t_{rise}) = L + \frac{H - L}{100} \longrightarrow t_0 - \frac{\log(100^{1/10} - 1)}{R} \quad (8.3)$$

### 8.3 Results

For each target-step condition, Figure 8.2 shows changes in saccade amplitude (y-axis) across adaptation trials (x-axis). In all plots, red and blue data show saccade amplitudes for the temporally and nasally moving eye, respectively, collapsed across subjects and eye movement direction. In good agreement with previous studies (Pélisson et al. 2010), in the outward step condition (Figure 8.2a), the saccades of both eyes increased in amplitude across adaptation trials; and in the inward step condition (Figure 8.2b), the saccades of both eyes decreased in amplitude. Critically, however, in the dichoptic step condition (Figure 8.2c), the saccades executed by the temporally moving eye (red curve) increased in size, while the saccades in the nasally moving eye (blue curve) decreased in size. Therefore, when presented with opposing error signals, the eyes adapt in different directions. It is worth noting however, that whereas in the early adaptation trials of the dichoptic step condition the saccade amplitudes indeed vary in different directions disjunctively, in the late adaptation the saccades executed by both temporally and nasally moving eyes appear to increase in size together, conjunctively.

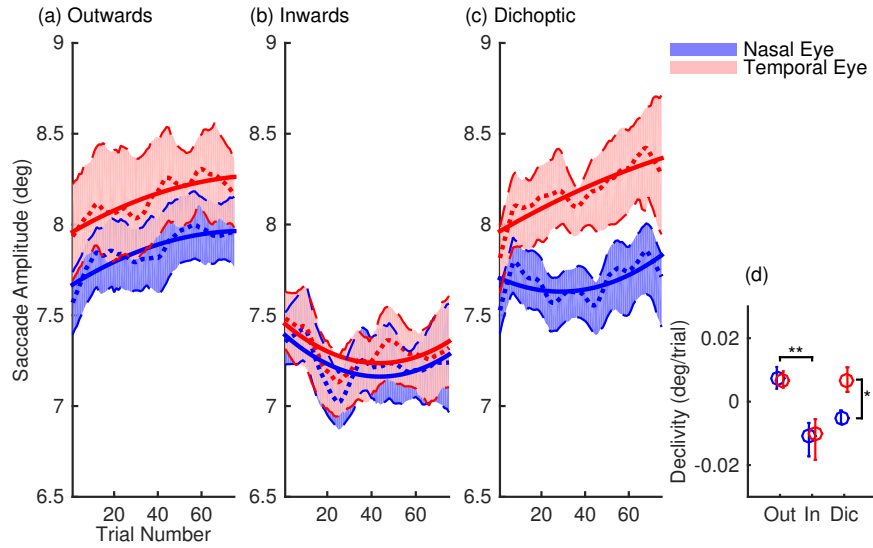


Figure 8.2: **Saccade adaptation.** (a-c) Saccade amplitude for temporally (red) and nasally (blue) moving eye as a function of trial number for outwards (a), inwards (b) and dichoptic (c) conditions. Data are smoothed with a lowess regression with a span of 15 trials and averaged across sessions and observers (dotted lines). Shaded regions represent 68% bootstrapped confidence intervals of the mean. Solid lines are the average polynomial equations fitted to the data (see Text for details). (d) Declivity parameter of the polynomial fits estimated for each eye and each condition, averaged across sessions and observers. Error bars are 68% bootstrapped confidence intervals. \*  $p < 0.05$ , \*\*  $p < 0.001$ .

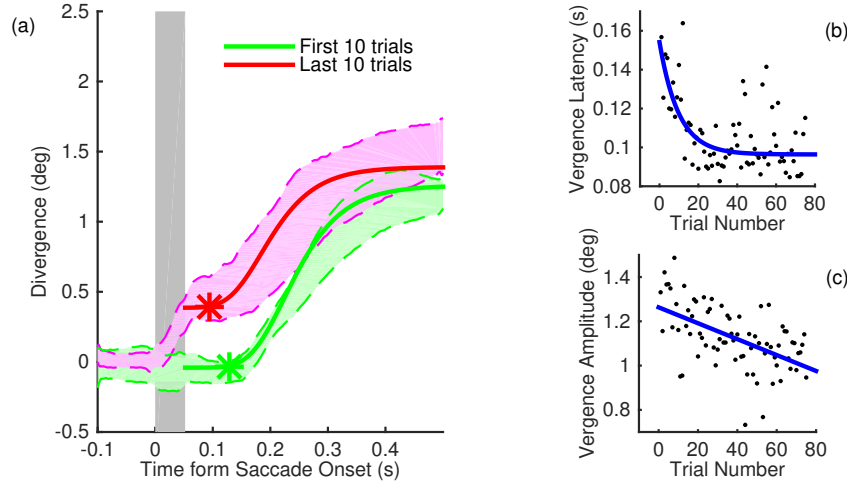
To quantify these observations, the saccade amplitude data were fit with second-degree polynomial (parabolic) equations, and quantitative estimates of the adaptation dynamics were extracted from these fits (see methods section 8.2). The initial gain of adaptation, given by the declivity parameter of the fit, is shown in Figure 8.2d. Note that the declivity parameter is independent of any baseline differences in saccade amplitudes across eyes and conditions. Adaptation gain increased for the outward step condition and decreased for the inward step condition (difference between outward and inward step conditions:  $p=0.000051$ ). In both these conditions, there was no significant difference in the adaptation gain between the temporally moving and nasally moving eyes (both  $p=0.99$ ). Conversely, in the dichoptic step condition, adaptation gain increased in the temporally moving eye and decreased in the nasally moving eye (difference between the temporally and nasally moving eye:  $p=0.036$ ). These data show that the saccades in each eye were independently recalibrated in opposite directions according to the different error signal simulated in each eye.

Immediately after each run of adaptation trials, observers completed a recovery procedure in which they performed the same saccade task but without the intra-saccadic target step. This was completed to observe the dynamics and ensure recovery from the transient eye deviations induced during adaptation. Note that relative to the adapted state of the observer, the change during recovery is inward in the temporal eye and outward in the nasal eye and thus examines the opposite direction of adaptation (convergent adaptation). Recovery patterns were the reverse of adaptation patterns for all three conditions, confirming that separable and reversible adaptation had occurred

in each eye.

In principle, the dichoptic adaptation described above could be attributable to independent monocular adaptation of the saccadic system for each eye, or binocular adaptation of vergence eye movements, as has been shown in the monkey (Schultz and Busetini 2013). To differentiate between these competing accounts, I examined inter-ocular divergence around the time of saccades (Figure 8.3). Figure 8.3a shows inter-ocular divergence (y-axis) as a function of time from saccade onset (x-axis) for the dichoptic step condition. During the first ten adaptation trials (green shaded region) all observers executed vergence eye movements following the saccades to foveate the dichoptic post-saccadic target. This vergence response (green curve) was initiated (green star)  $\sim 100$  ms after the saccade had ended. Were the changes in amplitude in the dichoptic step condition (Figure 8.2c) driven purely by vergence adaptation, this vergence function should shift leftward across trials. However, in contrast to the continuous vergence function for the first ten adaptation trials, the last ten adaptation trials (red shaded region) included two components. For all six subjects, first there was a rapid divergence (ranging from 0.1 degrees to 0.5 degrees across subjects) which occurred during the saccade (red shaded region overlapping the grey shaded region in Figure 8.3a), consistent with differential adaptation of saccade amplitudes in each eye. Only following this saccadic component is there a second vergence component, beginning after the end of the saccade (red star). These vergence eye movements also underwent adaptation, as revealed by their decreasing latency throughout the experiment (Figure 8.3b). Similarly, the amplitude of the post-saccade vergence response diminished throughout the adaptation

session (Figure 8.3c). Thus, vergence eye movements were modified during the dichoptic adaptation procedure, independently of saccadic amplitudes.



**Figure 8.3: Independent Vergence and Saccade Adaptation.** (a) Inter-ocular divergence around the time of a saccade following the dichoptic target step, averaged across observers and across the first 10 (green) and last 10 (red) trials. Grey shaded region represents the approximate period of the saccade. Red and green shaded regions are 68% bootstrapped confidence intervals of the median inter-ocular divergence as a function of time from saccade onset. Filled lines are the median logistic functions fitted to the data. Asterisks represent the point at which the vergence response is initiated. (b) Latency and (c) amplitude of the vergence response from saccade onset as a function of adaptation trial number. Data are the median for six observers. Solid blue lines are the best fitting exponential decay (latency) and linear (amplitude) functions passing through the data.

## 8.4 Discussion

In this chapter I have investigated monocular and binocular contributions to oculomotor plasticity. I replicated classic conjugate saccade adaptation results (McLaughlin 1967): an invisible intra-saccadic binocular target shift leads to an automatic adaptation of the saccade amplitude in both eyes. Using a dichoptic saccade paradigm I further showed that when the intra-saccadic displacement is in opposite directions in each eye (dichoptic step condition), adaptation occurs in opposite directions in each eye. Conversely, after removing the dichoptic step in the recovery trials, I found that the direction of adaptation reversed in both eyes, revealing convergent adaptation. Therefore, these data strongly suggest that our eyes receive at least partially independent recalibration commands. In the dichoptic step condition however, following an initial stage of apparently independent disconjugate adaptation one may also qualitatively observe some degree of conjugate adaptation, i.e. saccade amplitudes in both eyes appeared to increase in size together.

These results therefore inform the unresolved (King 2011) debate between Hering (1868) and Von Helmholtz (1867) regarding the nature of binocular oculomotor control. I show that binocular adaptation involves multiple potentially competing processes: independent disconjugate adaptation of monocular saccades, conjugate adaptation of binocular saccades, and binocular adaptation of vergence eye movements. Collectively, these adaptation processes likely serve to maintain binocular coordination while ensuring that the foveae of both eyes are directed to the intended locations of saccade targets. This plasticity could maintain saccade accuracy in the face of changes

in the forces acting upon the eye during normal development and aging (for a review see Salvi, Akhtar, and Currie 2006). These data also show that it may be possible to modify binocular coordination via oculomotor plasticity. This result informs our understanding of the development of normal binocular eye movement control in children (Fioravanti et al. 1995; Yang and Kapoula 2003) and has potentially significant implications for rehabilitation of oculomotor dysfunctions, including strabismus and vergence insufficiency.

Recent reviews of the literature have contrasted Helmholtz against Herring on theories of binocular eye movement coordination. These reviews have found evidence in favour and against both Herring's and Helmholtz's theories (King 2011) and suggest the controversy might be based on a false dichotomy (Coubard 2013). Premotor neurons in the paramedian pontine reticular have been found to encode monocular eye movements (Zhou and King 1998), yet it is possible that signals from these monocular neurons are combined downstream to produce binocular innervation of the eyes (King and Zhou 2002; Van Horn, Sylvestre, and Cullen 2008). Both monocular and binocular control networks likely coexist and cooperate to produce eye movements in depth. In this framework, recalibration mechanisms might act at multiple stages of the oculomotor control process and produce a complex pattern of adjustments. Independent adaptation of monocular saccades, conjugate adaptation of binocular saccades, and binocular adaptation of vergence could reflect different neural recalibration loci. This hypothesis is further supported by previous and apparently conflicting literature. Albano and Marrero (1995) have shown that adaptation induced only in one eye transfers to the other, occluded, eye, suggesting that recalibration was acting at the

level of binocular innervation. In the same study however Albano and Marrero also observed that when both eyes were unoccluded and an error signal was simulated in one eye only, adaptation only partially transferred to the other eye, a result I recently replicated (Maiello et al. 2014) and which suggest a combination of both monocular and binocular complimentary recalibration mechanisms. Independent monocular recalibration processes are also consistent with data from macaca fascicularis monkeys in which monocular surgical weakening of extraocular muscles produces rapid monocular recalibration without the involvement of vergence adaptation (Viirre, Cadera, and Vilis 1988). Conversely, Schultz and Busetini (2013) have recently observed binocular recalibration of the saccadic system via a mechanism which acts solely on the vergence system.

These results might also identify further differences in oculomotor control between human and other primates. The data presented in this chapter regarding classic conjugate saccade adaptation confirm that adaptation in humans can occur in tens of trials, whereas in monkeys adaptation takes several hundred trials (Hopp and Fuchs 2004). In rhesus monkeys, changes in saccade amplitude following dichoptic post-saccadic errors were argued to derive exclusively from vergence adaptation superimposed on a common saccadic command (Schultz and Busetini 2013). In humans, I find evidence of both rapid, independent monocular saccade adaptation together with a slower binocular vergence adaptation, consistent with evidence suggesting that humans have distributed neural adaptation loci yet monkeys have a common site of adaptation (Hopp and Fuchs 2004). It is however possible that monocular adaptation occurs in monkeys as well as humans, and that



the existing analyses, which focus on highlighting binocular vergence adaptation (see Figure 9 of Schultz and Busetini 2013), mask the existence of independent monocular saccade adaptation in monkeys.

This study highlights how understanding oculomotor recalibration in depth, which has scarcely been investigated in the literature, will help to map out and understand the neural control of binocular eye movements. Mapping out the resolution of oculomotor recalibration in 3D space (Chaturvedi and Van Gisbergen 1997), as well as separating out the individual saccadic and vergence (Takagi et al. 2001) based components of oculomotor recalibration, will allow for a more rigorous testing of more detailed models of eye movement control. The literature on eye movement control does not discuss the possibility of dissociable spatial maps for each eye, and it is generally assumed that a single distributed cortical representation of space is employed for the control of both eyes (T. J. Anderson et al. 1994; Corbetta 1998; Fox et al. 1985; O'Driscoll et al. 1995; Sweeney et al. 1996). The findings presented in this chapter raise the possibility that the oculomotor system maintains dissociable spatial maps for each eye or independently accesses a common map for both eyes.

## Chapter 9

# Three Dimensional Binocular Eye and Hand Coordination in Normal Vision and with Simulated Visual Impairments

### 9.1 Introduction

Humans have high resolution central vision and low resolution peripheral vision (Weymouth 1958). To visually explore the environment, humans move their eyes multiple times per second, rapidly shifting objects of interest selected with the low resolution peripheral retina onto the high resolution fovea (Dodge 1903). Thus, the oculomotor system continuously executes fast, ballistic *saccadic* eye movements (Javal 1879; Kowler 2011). In addition, to track moving objects the visual system executes slow, *smooth pursuit* eye

movements, which hold the fovea on a tracked object by matching its speed (Lisberger 2015). Furthermore, because humans have binocular overlapping visual fields, to inspect objects at different distances or to track objects moving in depth the oculomotor system also executes *vergence* eye movements, which are unequal slow rotations of the two eyes that shift the binocular gaze point in depth (Coubard 2013). The previous chapter investigated interactions between saccades and vergence eye movements. This chapter deals with how we visually track targets in 3D space, and thus I investigate interactions between smooth pursuit and vergence eye movements.

To actively interact with the environment, humans not only move their eyes, but also perform a range of complex hand movements which include reaching, grasping and manipulating objects. In order to take advantage of correlations that exist between spatial coordinates in visual and action space, the eye and hand motor systems have been found to be linked. Specifically, Steinbach and Held (1968) were the first to show that smooth pursuit eye movements to self-generated motion are more accurate than to externally generated motion, although Noorden and Mackensen (1962) had already shown that humans are able to make smooth pursuit eye movements to their own finger motion in the dark.

The ability to track the movement of one's digits without any visual information and the enhancement of oculomotor performance when tracking self-generated motion is likely to be due to an exchange of information between the eye and hand motor systems (Scarchilli and Vercher 1999). Proprioceptive information about hand location may also contribute (Mather and Lackner 1980, 1981). However, proprioception does not fully account for

increased accuracy at tracking self motion, since patients devoid of proprioception are still able to anticipate the start of smooth pursuit eye movements when tracking self-generated motion (Vercher et al. 1996). It is likely that the eye and hand system share motor planning information that facilitates the initiation of tracking eye movements, and then hand and arm proprioception helps maintain oculomotor tracking accuracy throughout the execution of the eye movements (Vercher and Gauthier 1992; Vercher, Quaccia, and Gauthier 1995; Vercher et al. 1996; Lazzari, Vercher, and Buizza 1997). A recent study (Chen, Valsecchi, and Gegenfurtner 2016) has shown that the lateralized readiness potential, which signals hand motor preparation, is associated with anticipatory smooth pursuit eye movements to self-generated finger motion, which is convincing evidence that the oculomotor system receives efference copy from the hand motor system. Not only is oculomotor performance enhanced by hand motor efference copy, but more broadly visual perception is enhanced near hand space (for a review see Perry, Amarasekera, and Fakhri 2016), suggesting some cross-modality overlap of spatial representations. Additionally, when visually tracking externally generated motion, manual tracking enhanced smooth pursuit eye movements, at least when the object motion is predictable (Gauthier et al. 1988; Koken and Erkelens 1992; Niehorster, Siu, and Li 2015). However, concurrent manual tracking has not been found to enhance vergence eye movements in depth (Koken and Erkelens 1993), and it is unknown whether the oculomotor system employs efferent hand motor information when tracking self-generated motion in depth.

Binocular vision and stereoscopic depth perception are fundamental in everyday eye-hand coordination (Fielder and Moseley 1996; O'Connor et

al. 2010; Gonzalez and Niechwiej-Szwedo 2016), and patients with stereovision deficits have impairments in reaching and grasping hand movements (e.g. in children with amblyopia, Grant et al. 2014, or in adults with age related macular degeneration ,Vergheze et al. 2016). Although hand-eye coordination deficits have been documented in these populations, one of the most basic and pervasive causes of visual impairment is refractive error, with an estimated 153 million visually impaired people worldwide as a result of uncorrected refractive errors (Resnikoff et al. 2008). Refractive error refers to the inability of the eye to focus light onto the retina. For example in chapters 5,6 and 7, we extensively discussed myopia, a common type of refractive error in which the eyes grow excessively large for the lens system to be able to focus far objects onto the fovea. Refractive error thus causes retinal images to be blurry, and blur reduces the contrast of mid to high spatial frequency image content, reducing the visibility of small objects and spatial detail. No official prevalence data is available, but for different age groups the distribution of myopic and hyperopic refractive errors is not negligible (Figure 9.1, Sorsby et al. 1960; Hyams, Pokotilo, and Shkurko 1977; Kamiya 1984; Fan et al. 2004). Some visually-impaired cohorts are defined by the type of refractive error. Patients with anisometropic amblyopia, for example, tend to have significant hyperopic refractive error that is asymmetric, which means that untreated amblyopes have different amounts of blur in the two eyes (Stewart et al. 2004).

It is currently unknown how eye-hand coupling is affected by reduced visibility. In conditions of visual uncertainty, the coupling between the eye and hand motor systems could break down. Conversely, if the oculomotor system

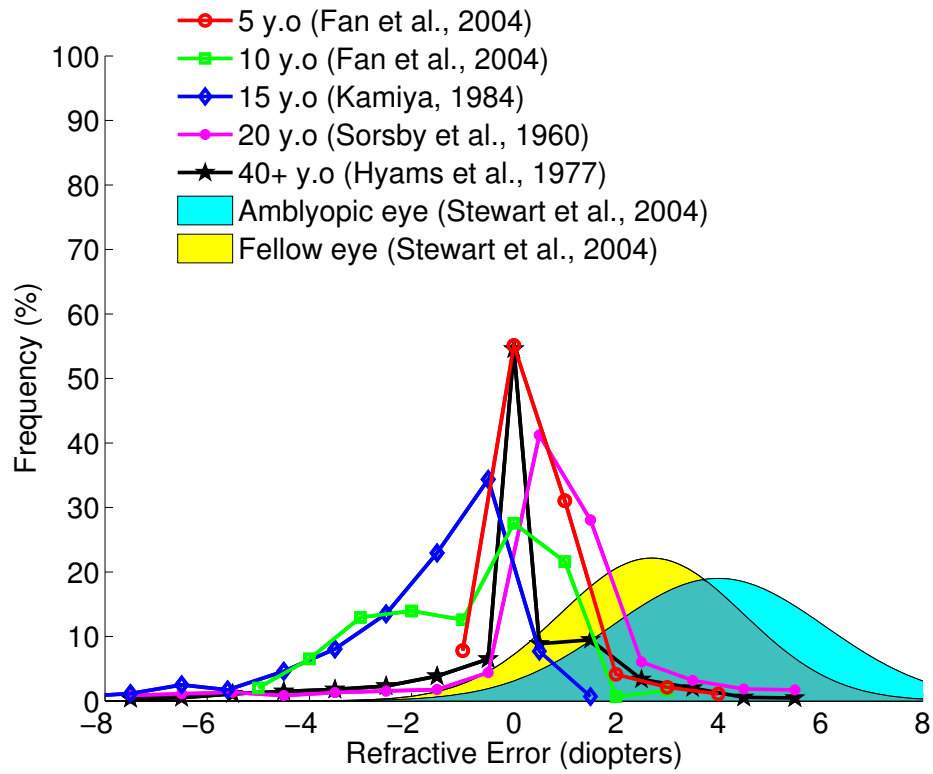


Figure 9.1: **Refractive error distributions.** Distributions of refractive errors for different age groups and for amblyopic children, replotted from a sample of the literature.

were optimally combining visual, hand motor efferent and proprioceptive information, it is possible that eye-hand coupling could be unaffected, or even strengthened, when visual information is unreliable. Thus, in the current chapter I investigate the coupling between the eye and hand motor systems under conditions of simulated monocular and binocular blur. I investigate both left-right pursuit eye movements in the fronto-parallel plane as well as vergence eye movements in depth. Because eye hand coordination occurs in 3D, I ask how eye-hand coupling in depth is affected by blur, and whether monocular and binocular visual impairments have similar effects. Lastly, because in this study blur is rendered, I discuss how the findings I present might relate to real-world blur due to refractive error.

## **9.2 Methods**

### **Participants**

Five subjects, author GM and four naïve observers, (3 male, mean  $\pm$ sd age:  $29 \pm 6$  ) participated in the study. All subjects had normal or corrected to normal vision and normal stereo vision. All subjects were right handed.

### **Apparatus**

The experiment was programmed with the Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) and Eyelink Toolbox (Cornelissen, Peters, and Palmer 2002) in Matlab (MathWorks). Stimuli were presented on an BenQ XL2720Z LCD monitor with a resolution of  $1920 \times 1080$  pixels (display

dot pitch 0.311 mm) at 120 Hz. The monitor was run from an NVidia Quadro K 420 graphics processing unit. Observers were seated 50 cm in front of the monitor with their heads stabilized in a chin and forehead rest and wore active wired stereoscopic shutter-glasses (NVIDIA 3D Vision) during all experiments to control dichoptic stimulus presentation. The cross talk of the dichoptic system was 1% measured with a Spectrascan 6500 photometer. Eye-tracking was performed using the EyeLink 1000 (SR Research) desktop mount eye-tracker running at 1000 Hz. The eye tracker was calibrated binocularly using the native five-point calibration routine at the start of each experimental session. Finger tracking was performed using a Leap Motion Controller, a commercial low-cost hand motion tracker recently validated for research applications (Weichert et al. 2013; Guna et al. 2014). Binocular gaze and finger position measurements were recorded at the monitor refresh rate of 120 Hz. The Leap Motion Controller was placed 30 cm in front and to the right of the observer, so that the observer’s right hand could be tracked.

## Stimuli

The stimulus for target tracking was a Gabor patch ( $\sigma = 0.25$  degrees,  $\omega = 2$  cycles/degree, 100% Michelson contrast) moving on top of a white bar (0.5 degree thick and 8 degree wide). Stimuli were embedded in 1/f pink noise background which has the same frequency content of natural images (Kretzmer 1952; Bex and Makous 2002) and helps reduce the visibility of stereoscopic crosstalk. Example stimuli are shown in Figure 9.2. From trial to trial the amount of blur in the stimulus to one or both eyes was systematically varied using fast Gaussian mipmap filtering. Blur level is specified as double



the standard deviation of the Gaussian blur kernel in minutes of arc. In the pursuit sessions seven blur levels were employed: [0, 6.25, 12.5, 25, 50, 100, 200] arcmin. In the vergence sessions only 6 blur levels were employed, as in pilot testing the vergence task was found to be already near impossible with 100 arcmin of blur. It is important to note that with increasing blur levels the contrast of the stimuli decreased. The blur levels employed can be equated to contrast levels of [100, 95, 78, 59, 48, 35, 20] % Michelson contrast.

In the binocular blur trials, the stimuli shown to both eyes were blurred. In the monocular blur trials instead, only the stimuli shown to the left eye were blurred, whereas the right eye was always shown a sharply focused stimulus.

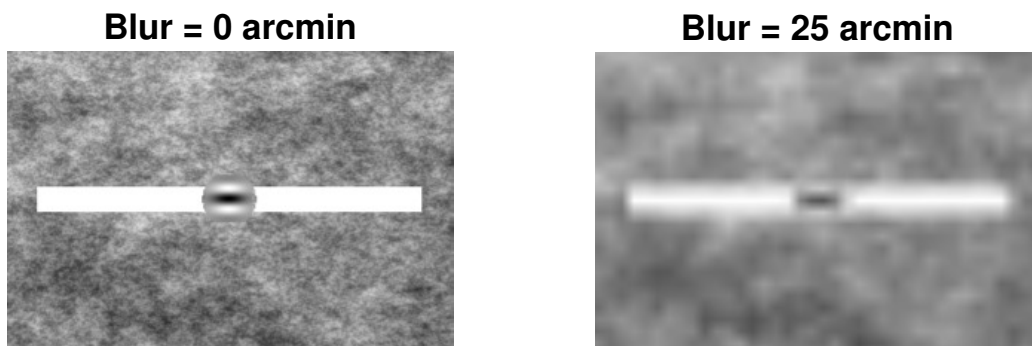


Figure 9.2: **Example stimuli for target tracking.** The stimulus for target tracking was a gabor patch moving on top of a white bar. Background was 1/f pink noise. Left panel is an example stimulus frame in the 0 arcmin blur condition. Right panel is an example stimulus frame in the 20 arcmin blur condition.

## Design and Procedure

Observers participated in four experimental sessions on separate days. The four sessions are schematized in Figure 9.3. In sessions 1 and 2, which were horizontal pursuit eye movement sessions, observers were required to complete 5 trials for each of 7 blur levels for both monocular and binocular blur conditions. Thus in sessions 1 and 2 observers completed 70 trials per session. In sessions 3 and 4, which were vergence pursuit eye movement sessions, observers completed 5 trials for each of 6 blur levels for both monocular and binocular blur conditions, for a total of 60 trials per session. Within each session trial order was fully randomized.

- Session 1: Horizontal Pursuit Finger Tracking (Figure 9.3a). Each trial, observers were required to smoothly move their index finger from right to left and back, directly above the Leap Motion Controller, which tracked the motion of the observer's fingertip. The on-screen Gabor target moved with the observer's finger, and the observer was required to track the Gabor target as accurately as possible with his/her gaze.
- Session 2: Horizontal Pursuit Replay Tracking (Figure 9.3b). Observers were required to hold their hands motionless. Each trial, the Gabor target moved right to left and back, replaying one of the finger movements executed in session 1, matched to the viewing condition. The observer's task was solely to track the Gabor target as accurately as possible with his/her gaze.
- Session 3: Vergence Pursuit Finger Tracking (Figure 9.3c). Each trial, observers were required to smoothly move their index finger backwards

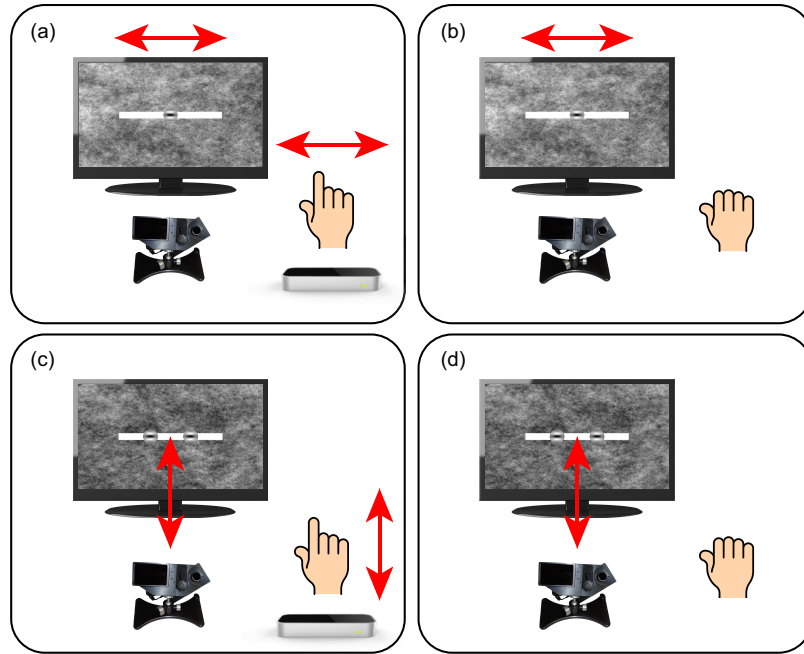


Figure 9.3: **Schematics of the four experimental target tracking sessions.** In all four sessions observers' binocular gaze point was monitored with an Eyelink eyetracker. (a) In session 1 observers smoothly moved their right index finger from right to left and back, directly above the Leap Motion Controller. The on-screen Gabor target moved with the observers' finger. (b) In session 2 observers' hands were motionless and the Gabor target moved right to left and back, replaying one of the finger movements executed in session 1. (c) In session 3 observers smoothly moved their right index finger backwards and forwards, directly above the Leap Motion Controller. The on-screen Gabor target moved in stereoscopic depth with the observers' finger. (d) In session 4 observers' hands were motionless and the gabor target moved backwards and forwards, replaying one of the finger movements executed in session 3.

and forwards, directly above the Leap Motion Controller, which tracked the motion of the observer’s fingertip in depth. The on-screen Gabor target moved in depth with the observer’s finger, and the observer was required to track (by executing vergence eye movements) the Gabor target as accurately as possible with his/her gaze.

- Session 4: Vergence Pursuit Replay Tracking (Figure 9.3d). Observers were required to hold their hands motionless. Each trial, the Gabor target moved backwards and forwards, replaying one of the finger movements in depth executed in session 3, matched to the viewing condition. The observer’s task was solely to track the Gabor target as accurately as possible with his/her gaze.

## Analysis

Each trial, tracking accuracy was computed as the correlation between gaze position and target position either along the fronto-parallel plane (horizontal pursuit) or the sagittal plane (vergence pursuit). Figure 9.4 shows tracking accuracy as a function of blur for a representative observer in the binocular blur, horizontal pursuit finger tracking condition. These data were first transformed via Fisher’s Z transformation to ensure variance stabilization (Fisher 1915). Correlation data were then fit to a hinged line linear-log function with equation:

$$y = \begin{cases} b & : \log_2(x) < c \\ d \log_2(x) + b - dc & : \log_2(x) \geq c \end{cases} \quad (9.1)$$

where  $b$  is an observer’s *baseline* level of tracking accuracy,  $c$  is the *critical*

*blur* level at which tracking performance begins to deteriorate, and  $d$  is the rate of *decay* of tracking accuracy with increasing levels of blur beyond  $c$ . Parameter estimates were analyzed using a  $2$  (Finger vs Replay Tracking)  $\times$   $2$  (Pursuit vs Vergence)  $\times$   $2$  (Binocular vs Monocular blur ) within-subject ANOVA.

### 9.3 Results

Human observers performed a simple target tracking task. A Gabor target moved either left-right (horizontal pursuit tracking condition) or backwards and forwards in depth (vergence pursuit tracking condition). Observers were required to track the target as accurately as possible with their gaze. In two experimental sessions, the movement of the target was controlled by the movements of the observer’s index finger (finger tracking condition), whereas in two experimental sessions observers stilled their hands and the target moved according to previously recorded finger movements (replay tracking condition). From trial to trial, I varied the visibility of the target by blurring the scene with varying amounts of blur, either to both eyes (binocular blur condition) or solely to the left eye (monocular blur condition). Tracking accuracy was estimated as the correlation between gaze position and target position either along the fronto-parallel plane (horizontal) or the sagittal plane (vergence). Correlation data as a function of blur level were first scaled according to Fisher’s Z transform to ensure variance stabilization (Fisher 1915) and then fit to hinged line linear-log functions that provided estimates of the *baseline* level of accuracy (i.e. accuracy when no blur was present), the

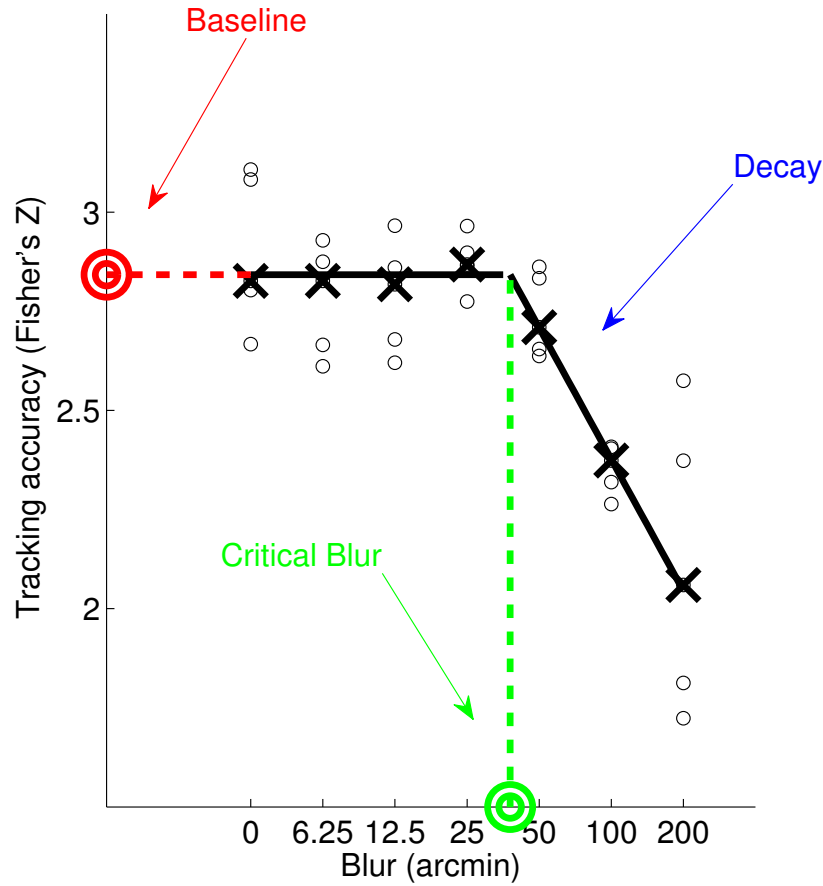


Figure 9.4: **Example raw accuracy data and hinged line fitting.** Tracking accuracy as a function of blur for a representative observer in the binocular blur, finger tracking, horizontal pursuit condition. Black circles are individual trial accuracy data. Black crosses are medians. Black lines are fitted hinged line linear-log functions. The y-axis intercept is the *baseline* level of accuracy (red bullseye). Beyond the *critical blur* level (green bullseye), accuracy begins to falls below the baseline level of tracking. The slope of the linear falloff (indicated by the blue arrow) is the rate of *decay* of tracking accuracy with increasing levels of blur.

*critical blur* level at which performance began to deteriorate, and the rate of *decay* of tracking performance beyond the *critical level* of blur.

Tacking accuracy varied lawfully as a function of blur level (Figure 9.5a,b); accuracy was at a constant *baseline* up to a *critical blur* level, after which accuracy fell-off on linear-log axes with a linear rate of *decay*. This pattern held true for both horizontal pursuit eye movements (Figure 9.5a,b squares) and vergence pursuit eye movements (Figure 9.5a,b triangles).

Tracking accuracy was higher when observers tracked a stimulus controlled by their own finger movements in real time (black curves, Figure 9.5a,b) than when observers tracked a replay of their own previously executed finger movements (red curves, Figure 9.5a,b). Specifically, *baseline* accuracy was greater in the finger tracking than the replay condition (Finger vs Replay main effect:  $F(1,4) = 19.96$ ,  $p = 0.011$ ). Horizontal pursuit tracking accuracy was three-fold greater (in Z-scaled space) than vergence pursuit tracking accuracy (Horizontal vs Vergence main effect on *baseline* accuracy:  $F(1,4) = 153.82$ ,  $p = 0.00024$ ). However, the difference between finger and replay conditions was similar whether tracking was horizontal/left-right or in depth (Interaction effect on *baseline* accuracy between Finger vs Replay and Horizontal vs Vergence:  $F(1,4) = 0.080$ ,  $p = 0.79$ ). This result further confirms the coupling between the oculomotor and hand motor systems and shows that the eye and hand motor systems are also linked for movements in 3D space. Sensibly, *baseline* accuracy did not vary as a function of monocular or binocular blur conditions, as these conditions are the same at baseline for the least blurred condition (Monocular vs Binocular blur main effect on *baseline* accuracy:  $F(1,4) = 0.078$ ,  $p = 0.79$ ; all two and three way interactions:

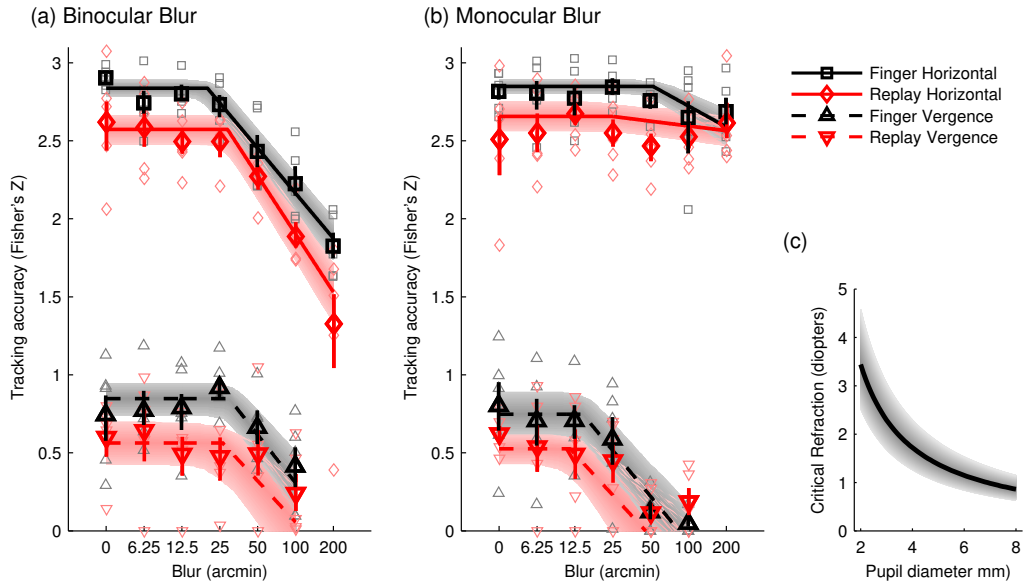


Figure 9.5: **Tracking accuracy, blur and refraction.** Tracking accuracy at each blur level for binocular (a) and monocular (b) blur conditions. Squares are data for horizontal pursuit eye movements, triangles are data for vergence pursuit eye movements. Finger tracking data is in black, replay tracking is in red. Large symbols are average tracking accuracy across subjects at each blur level. Error bars are 68% confidence intervals. Small symbols are individual subject data. Filled and dashed lines are average fitted hinged line linear-log functions for horizontal pursuit and vergence pursuit respectively. Shaded regions are 68% confidence regions of the fitted functions. Note that the y-axis is scaled following Fisher's Z transformation, whereas the x-axis is log scaled. (c) The estimated Critical Refraction at which oculomotor tracking performance may begin to decay is plotted as a function of pupil diameter (black curve). Shaded region encompasses the 95% confidence range of the estimate.



$p > 0.16$ ).

Tracking accuracy remained higher in the finger tracking condition compared to the replay tracking condition even when blur disrupted tracking performance. The rate of *decay* of tracking accuracy beyond the *critical blur* level did not vary between finger and replay tracking conditions (Finger vs Replay main effect:  $F(1,4) = 0.53$ ,  $p = 0.51$ ). The rate of *decay* of tracking accuracy was also similar for horizontal and vergence eye movements (Horizontal vs Vergence main effect:  $F(1,4) = 1.23$ ,  $p = 0.33$ ). The rate of *decay* did differ across monocular and binocular tracking conditions (Monocular vs Binocular blur main effect:  $F(1,4) = 43.04$ ,  $p = 0.0028$ ), and more specifically as a function of whether tracking was left-right or in depth (Interaction effect between Monocular vs Binocular blur and Horizontal vs Vergence:  $F(1,4) = 9.87$ ,  $p = 0.035$ ). When observers executed pursuit eye movements, the rate of *decay* was steep if blur was rendered to both eyes and shallow if blur was rendered to one eye only (compare horizontal pursuit data in Figures 9.5a and b), and this difference was statistically significant ( $p=0.00052$ , paired samples t-test). When observers executed vergence eye movements instead, the rate of decay was similarly steep in both monocular and binocular blur conditions (compare vergence pursuit data in Figures 9.5a and b;  $p=0.38$ , paired samples t-test). All other ANOVA two and three way interactions were not statistically significant (all  $p > 0.19$ ).

The overall *critical blur* level at which tracking accuracy began to fall off was found to be 19 arcmin [14-25 arcmin, 95% confidence range]. This *critical blur* value was independent of the type of eye movement executed (Horizontal vs Vergence main effect:  $F(1,4) = 1.41$ ,  $p = 0.30$ ), of whether

observers were tracking their own finger movements or a replay of previously executed finger movements (Finger vs Replay main effect:  $F(1,4) = 0.81$ ,  $p = 0.42$ ) and of whether blur was rendered monocularly or binocularly (Monocular vs Binocular blur main effect:  $F(1,4) = 1.031$ ,  $p = 0.37$ ; all two and three way interactions:  $p > 0.22$ ).

A very simple approximate relationship exists between blur and refraction (Smith 1982):

$$b = p_d R \quad (9.2)$$

meaning that defocus in angular units  $b$  (radians) is equal to the refractive status  $R$  (diopters) times the pupil diameter  $p_d$  (meters). Given that we have estimated the *critical blur* level at which tracking performance begins to deteriorate, we can employ Equation 9.2 to estimate at what level of refractive error oculomotor tracking performance begins to break down. Figure 9.5c shows this estimated Critical Refraction as a function of pupil diameter. Pupil size is strongly dependent on light levels: pupils will constrict in high light and enlarge with low light levels. Thus, for high, outdoor light levels, up to 3 diopters of refractive error may be necessary to hinder oculomotor performance. For low, indoor light levels instead, 1 diopter of refractive error may already be sufficient to degrade oculomotor tracking performance.

## 9.4 Discussion

I investigated the link between the oculomotor and hand motor control systems under conditions of simulated asymmetric visual blur, for both hori-

zontal pursuit eye movements and vergence pursuit eye movements in depth. I replicated the classic result that smooth pursuit eye movements to self-generated horizontal motion are more accurate than to externally generated motion (Steinbach and Held 1968). I extend these classic studies to show that the link between the eye and hand motor systems also exists for eye and hand movements in depth. Additionally, I show that moderate levels of simulated visual blur disrupts oculomotor tracking, but the link between the eye and hand motor system persists at all blur levels. Lastly, monocular and binocular blur have different effects on eye movements in the fronto-parallel (horizontal pursuit eye movements) and saggital (vergence pursuit eye movements in depth) planes. Binocular blur similarly disrupts both horizontal pursuit and vergence pursuit eye movements. Monocular blur instead strongly affects vergence eye movements, but only mildly hinders horizontal pursuit eye movements.

Concurrent hand tracking aids smooth pursuit tracking performance of externally-moving stimuli (Gauthier et al. 1988; Niehorster, Siu, and Li 2015) but only when the motion is predictable (Koken and Erkelens 1992). Koken and Erkelens (1993) have further shown that when observers execute vergence eye movements to track a stimulus moving predictably in depth under external control, concurrent hand tracking of the stimulus did not aid oculomotor tracking performance. These result suggests that the oculomotor system may employ limb proprioceptive (afferent) signals to plan eye movements in the fronto-parallel plane, but is unable to do so for eye movements in depth. Conversely, I find that tracking accuracy is higher for both horizontal pursuit and vergence eye movements when observers track self-generated

hand motion. Thus, the efferent signals from the hand motor system that are employed by the oculomotor system to facilitate the initiation of horizontal smooth pursuit eye movements (Chen, Valsecchi, and Gegenfurtner 2016) are also likely employed in planning the initiation of vergence tracking eye movements in depth.

If the oculomotor system were able to modulate the contributions of the visual input and the hand motor efferent signal when planning horizontal smooth pursuit and vergence tracking eye movements, we might expect the hand motor efferent signal to be weighed more strongly in conditions of greater visual uncertainty. However, the rate of *decay* of tracking accuracy was the same in the finger and replay tracking conditions. Thus, these data suggest that the contribution of the hand motor efferent signal to oculomotor planning may be fixed and invariant over the short times covered in the present study. These results cannot rule out changes in the contribution of motor efferent signals over longer periods of adaptation to visual impairment.

The *critical blur* level at which tracking accuracy began to decay was independent of all experimental manipulations in the present study. I estimated that between 1 and 3 diopters of refractive error may be sufficient to produce this level of critical blur, depending on the ambient light level. Consider this *critical refraction* with respect to the distribution of refractive errors found in the general population (Figure 9.1). A significant portion of the population has refractive errors that fall beyond the estimated *critical refraction* level. Furthermore, amblyopic subjects and other patients with asymmetric or unilateral visual impairment (such as cataracts), not only have an average acuity loss beyond this level of *critical blur*, but even the average

difference in refraction between the amblyopic and fellow eye (1.3 diopters) might be enough to hinder oculomotor performance. Furthermore, I find that monocular blur hinders vergence but not left-right eye movements. This is sensible, since for fronto-parallel tracking one eye is sufficient to track the stimulus, while the other eye can simply follow. However, detailed vision in both eyes is necessary to estimate binocular disparity to drive vergence eye movements. Asymmetric refraction is thus likely to affect some oculomotor tasks more than others. Reading (Rubin et al. 2000) or viewing content on a flat monitor might be less hindered by monocular visual impairment than, for example, eye-hand coordination tasks such as reaching and grasping.

It is important to note that in this study visual impairments were simulated by Gaussian blurring the visual input. It is possible that oculomotor performance may differ with different kinds of blur (such as sinc blur which contains phase reversals typical of the modulation transfer function of an optical system with a circular aperture such as the human pupil, Murray and Bex 2010), and that subjects may learn to adapt to the specific type of defocus blur arising from their own optics (Artal et al. 2004). Future investigations should thus attempt to study hand-eye coordination and oculomotor control in the general population as well as monocular visual acuity for static optotypes. Cross-sectional studies spanning the range of refractive errors present in the general population may help identify the critical refraction at which oculomotor control begins to break down in observers who have adapted to their visual deficits. Longitudinal studies in children may help identify critical time periods in which refractive errors may hinder visual and oculomotor development.

These findings may also inform the aetiology of strabismic ambliopia. It is thought that this visual dysfunction develops when binocular vision becomes decorrelated owing to alignment or refractive error in one eye that causes the brain to suppress visual information from this deviated eye to avoid diplopia (double vision) (Wright, Spiegel, and Thompson 2006). However the development of oculomotor and perceptual deficits might be interdependent. If the visual input from one eye is not correlated with the other eye, oculomotor control might begin to degrade synergistically. If this occurs during early life, when vision and oculomotor control is in development (Atkinson, Braddick, and Moar 1977; Banks and Salapatek 1978; Roucoux, Culee, and Roucoux 1983; Hainline et al. 1984; Hainline and Riddell 1995), poor oculomotor control due to unbalanced visual input might form part of a dysfunctional system that exacerbates incorrect eye posture (ocular deviation). This deviation might in turn lead to suppression of visual input from the already weak eye, causing a feedback loop of visuomotor impairment.

This work also has potential implications for visual rehabilitation strategies in which binocular eye movements are impaired, such as strabismus and convergence insufficiency. Classical patching therapy might be counterproductive for the development of coordinated vergence eye movements because it forces the two eyes to work independently and does not favour the correct development of binocular oculomotor control. Dichoptic therapies that attempt to balance the input to the two eyes might be better suited to ensure binocular cooperation (Hess, Mansouri, and Thompson 2010; To et al. 2011; Knox et al. 2012; J. Li et al. 2013; Tailor et al. 2015; Tailor et al. 2016). I have shown that the link between the oculomotor and hand motor sys-

tem is unaffected by simulated visual impairment. Thus, eye-hand coordination tasks (Vedamurthy et al. 2016) might be able to provide a boost to visual rehabilitation strategies by enhancing oculomotor control both along the fronto-parallel plane and particularly in depth.

## Chapter 10

# A Space-variant Model for Motion Interpretation across the Visual Field

### 10.1 Introduction

In this chapter I examine the role of central and peripheral vision in the dorsal visual pathway (Goodale and Westwood 2004; Orban 2008), for the processing of self-motion and object motion (Watson, Ahumada Jr, et al. 1985; Koenderink 1986; Orban et al. 1992; Perrone and Stone 1994; Simoncelli and Heeger 1998; Rushton and Warren 2005; Warren and Rushton 2007, 2009). More specifically, I examine an existing model that is able to compute the focus of expansion (FOE), which is the point from which all motion vectors expand, indicating the direction of heading when eye and head positions are known. Humans can accurately judge direction of heading by observing the



location of the FOE in the flow field (Warren Jr and Hannon 1988). The focus of radial motion (FRM) is a generalization of the FOE: it is the point from which all motion vectors either expand or contract. Bex and Falkenberg (2006) showed that the accuracy with which human observers can estimate the FRM in random dot stimuli decreases as a function of retinal eccentricity and, using equivalent noise analysis, that this effect may be due to an increase in local motion detector noise rather than undersampling.

I study a neural model which implements the deep hierarchy of the first stages of the dorsal visual pathway (Solari, Chessa, and Sabatini 2014). Such a model (see Figure 10.1) is space-variant, since it takes into account the retino-cortical transformation of the primate visual system through log-polar mapping that produces a cortical representation of the visual signal to the retina. I study how this space-variant approach affects optic flow computations across the visual field. Specifically, I ask whether a model which implements the known processing stages of the dorsal visual pathway is sufficient to account for human perception. In particular, in this chapter (i) I describe the neural model for the estimation of the FRM at different retinal locations, and (ii) I assess the model by comparing its results with respect to (iii) the experimental evidence related to the precision with which human observers can estimate the FRM in naturalistic, dead leaves stimuli (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001).

The first stage of the proposed model implements the retino-cortical transformation by mapping the cartesian retinal image onto log-polar cortical coordinates. The log-polar transform of the retinal image is the input to the cortical motion estimation stage, where optical flow is computed. The

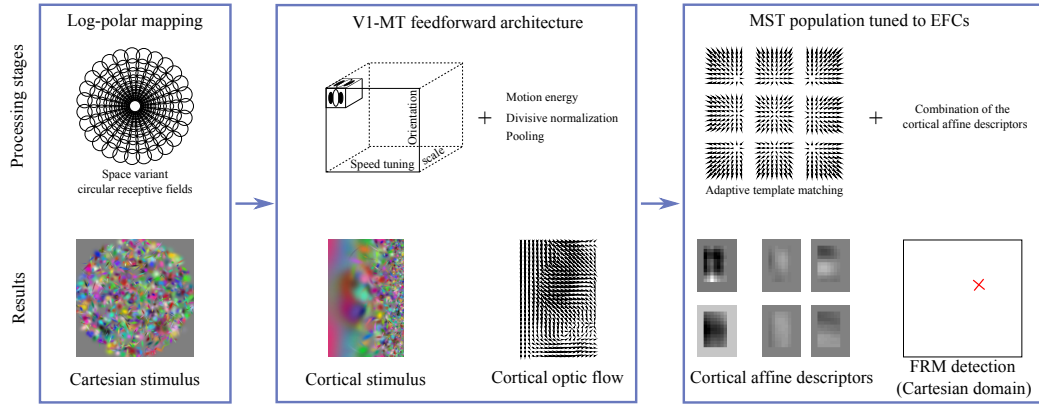


Figure 10.1: **The neural space-variant model.** The Cartesian stimulus (i.e. the sequence of dead leaves) is transformed into the cortical domain through the log-polar mapping, then a V1-MT feedforward architecture produces an estimation of the cortical optic flow, on which a population of MST-like cells tuned to Elementary Flow Components (EFCs) is used to estimate the affine (i.e. first-order description) of the cortical optic flow. A combination of such first-order descriptors produces an estimation of the FRM in the Cartesian domain.

cortical motion estimation stage consists of a three-layer population of cells. A population of spatio-temporal oriented Gabor filters approximates the simple cells of area V1 (first layer), which are combined into complex cells as motion energy units (second layer). The responses of the complex cells are pooled (third layer) to encode the magnitude and direction of velocities, as in the extrastriate motion pathway, between area MT and MST. The sensitivity to complex motion patterns that has been found in area MST is modeled through a population of adaptive templates, and from the responses of such a population the first order description of optic flow is derived. Information about self-motion (e.g. direction of heading) is estimated by combining such first-order descriptors computed in the cortical domain.

I thus investigate whether a fully functional, space variant model of the known processing stages of the dorsal visual pathway accounts for human performance across the visual field. Such a model must take real image sequences as input, perform space variant computations on the input images, and output behaviorally relevant metrics such as the estimate of the FRM. From this investigation I conclude that the model shows qualitatively and quantitatively similar patterns of results to the human behavioral data. The proposed neural model thus captures essential aspects of the neural computations that occur in the cortical motion pathway. I also discuss which aspects of human perception are not captured by the model and might be interesting starting points for further investigation. This work has clear applications in neuromimetic robotic architectures and, more broadly, it provide a framework in which to model complex motion integration across the visual field.

**Related work** From neurophysiological studies it is well known that Neurons in the dorsal part of the medial superior temporal area (MSTd) are selective for optic flow patterns (Tanaka and Saito 1989) generated by the relative motion between the observer and the 3D world. MSTd neurons are known to respond to elementary (expansion, contraction, rotation, and translation) as well as complex patterns of optic flow (Duffy and Wurtz 1991; Graziano, Andersen, and Snowden 1994). It has been shown that activity in MSTd neurons that are selective for expanding optic flow fields is linked with behavioral performance in heading judgments (Gu et al. 2010; Gu, DeAngelis, and Angelaki 2012). Xu, Wallisch, and Bradley (2014) conclude that a variety of subpopulations of MSTd neurons, differently tuned for expansion, rotation or spiral motion, contribute to heading judgments. Graziano, Andersen, and Snowden (1994) analyzed the idea that navigation is achieved by decomposing the optical flow into the above-mentioned elementary components. They found that many cells in MST are tuned to intermediate spiral motions, obtained by the combination of expansion and rotation components. These findings reinforced the idea that MST might process complex configurations of visual motion, to obtain information both on navigation and on motion of objects and surfaces. In the human brain, the analysis of visual motion continues in higher processing layers. Duffy and Wurtz (1995) found that many neurons in MST area respond to a combination of expansion and planar motions which shifts the position of the FOE in the neuron’s receptive field. In particular, they found a continuum in the MST neurons’ responses, which combine expansion, rotation and planar motions. Recently Xu, Wallisch, and Bradley (2014) described the *Spiral space model*: they propose

that neurons in MSTd are tuned to elementary optic flow types (and consequently to motion in real-world space), in a continuous way, which may be called “spiral space” tuning. In particular, they propose that adding laminar motion (e.g., a translation to the left) to radial motion (e.g., expansion) yields tuning for a particular FRM location or direction of heading in real situation. Their findings suggest that MSTd neurons respond to all possible motion types in a 3D space (called the spiral space) whose axes are the expansion, rotation and laminar motion components (see Figure 1 in Xu, Wallisch, and Bradley 2014).

This rich array of neurophysiological findings has led to the development of multiple computational models of optic flow processing and heading perception. Lappe and Rauschecker (1993) propose a neural model that detects the direction of ego-motion from optic flow in a fashion that is consistent with neurophysiological and psychophysical data. The model however does not implement optic flow estimation, but describes the neural computations occurring after layer MT. A “proof of principle” of a template-like strategy for heading estimation by following a strategy similar to MSTd neurons is presented in Perrone and Stone (1994). The input units of such a model are idealized MT neurons. Grossberg, Mingolla, and Pack (1999) describe a neural model that provides both functional explanation and quantitative simulation of experimental data of MSTd cells. Grossberg and colleagues interestingly take into account the retino-cortical mapping by providing log-polar representations of optic flow patterns as inputs to their model. Beardsley and Vaina (2001) describe a model population of MST neurons capable of performing graded motion pattern discrimination tasks. More recently, the ViS-

TARS neural model proposed in (Browning, Grossberg, and Mingolla 2009) describes interaction among neurons in the primate magnocellular pathway and is capable of estimating heading direction from natural image sequences. The ViSTARS model however does not consider the retino-cortical transformation, and is thus space-invariant. Previous work directly comparing the performance of humans and computational models at estimating heading direction from optic flow (Foulkes, Rushton, and Warren 2013b) has shown that existing models capture some, but not all, aspects of human behaviour.

The model I investigate here incorporates known aspects of neural processing relevant for motion estimation and interpretation. The retinal image input to the model is mapped onto cortical space where optic flow is estimated and decoded directly in cortical coordinates to provide various behaviorally relevant metrics, including direction of heading.

## Model Description

The biologically-inspired neural model of the dorsal motion pathway, originally developed and described by Solari, Chessa, and Sabatini (2014) can be summarized as follows.

**Retino-cortical mapping** Each Cartesian moving image sequence is transformed into its cortical representation through a log-polar transformation that implements the central blind-spot model (CBSM). Such a transformation mimics the *retino-cortical mapping* of the primate visual system (Schwartz 1977; Traver and Pla 2008; Wilkinson et al. 2016).

In the CBSM, the mapping from the Cartesian domain  $(x, y)$  to the cor-

tical domain of coordinates  $(\xi, \eta)$  is described by the following equations:

$$\begin{cases} \xi = \log_a \left( \frac{\rho}{\rho_0} \right) \\ \eta = q\psi, \end{cases} \quad (10.1)$$

where  $a$  parametrizes the non-linearity of the mapping,  $q$  is related to the angular resolution,  $\rho_0$  is the radius of the central blind spot, and  $(\rho, \psi) = (\sqrt{x^2 + y^2}, \arctan(y/x))$  are the polar coordinates derived from the Cartesian ones. All points with  $\rho < \rho_0$  are ignored (hence the central blind spot), thus  $\rho_0$  must be small, with respect to the size of the image.

*Discrete log-polar mapping.* Because the transformation is implemented on digital images, given a Cartesian image of  $N_c \times N_r$  pixels, and defined  $\rho_{max} = 0.5 \min(N_c, N_r)$ , we obtain an  $R \times S$  (rings  $\times$  sectors) discrete cortical image of coordinates  $(u, v)$  by taking:

$$\begin{cases} u = \lfloor \xi \rfloor \\ v = \lfloor \eta \rfloor, \end{cases} \quad (10.2)$$

where  $\lfloor \cdot \rfloor$  denotes the integer part,  $q = S/(2\pi)$ , and  $a = \exp(\ln(\rho_{max}/\rho_0)/R)$ . Moreover, we can define the compression ratio ( $CR$ ) of the cortical image with respect to the Cartesian one as:

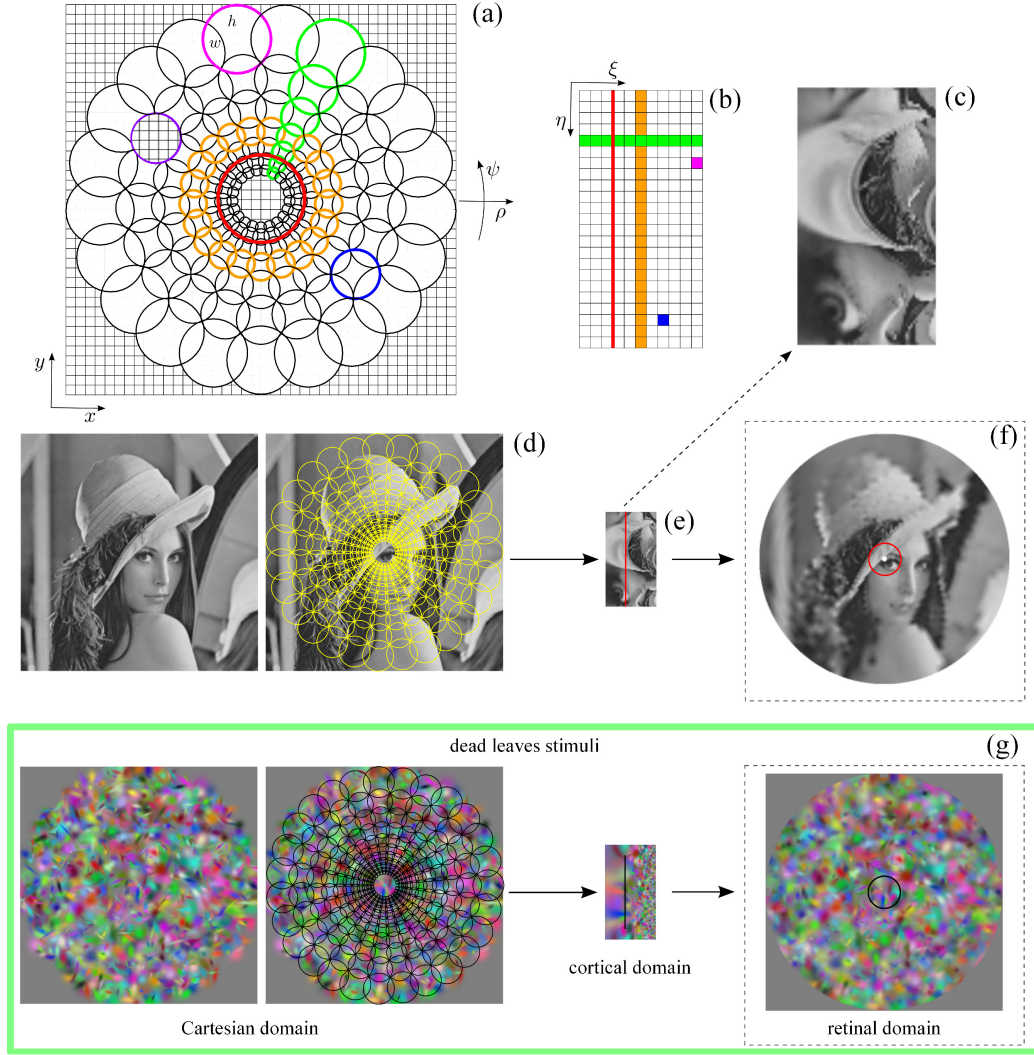
$$CR = \frac{N_c \times N_r}{R \times S} \quad (10.3)$$

Figure 10.2 shows the log-polar pixels (which can be thought of as the log-polar receptive fields) in the Cartesian domain (a) and in the cortical domain (b). The red circular curve in (a), with radius  $S/2\pi$ , represents the locus where the size of log-polar pixels is equal to the size of Cartesian pixels. In particular, in the area inside the red circular curve (the *fovea*) a single

Cartesian pixel contributes to many log-polar pixels (oversampling), whereas outside this region multiple Cartesian pixels will contribute to a single log-polar pixel. This property is highlighted in the log-polar pixel bordered in violet in (a). The retinal area (i.e. the log-polar pixel) that refers to a given cortical pixel defines the cortical pixel’s *receptive field* (RF). To avoid spatial aliasing due to the undersampling, overlapping RFs are employed. An example of transformation from the Cartesian domain to the cortical domain and back to the retinal one is shown in Figure 10.2, for the standard image *Lena* (c-d-e-f) and for a frame (g) of the dead leaves stimuli used in the experiments described below. It is worth noting that the cortical image (e) shows the effects of the log-polar mapping: in particular, the zoomed cortical image (c) shows that the eye, which is in fovea in (d), is over-represented (left-bottom of (c)), whereas the hat feathers, which are in periphery in (d), is under-represented (right-middle of (c)), i.e. few neural units code the visual information of the periphery. By looking at the backward mapped image (f), the eye has full resolution, since it is in fovea, whereas the hat feathers have low resolution due to the neural under-representation.

By inverting Equation 10.1 the centers of the RFs can be computed, and these points present a non-uniform distribution throughout the retinal plane (see the yellow and black circles overlying the Cartesian images of *Lena* and dead leaves, respectively, in Figure 10.2(d-g)). The optimal relationship between  $R$  and  $S$  is the one that optimizes the log-polar pixel aspect ratio  $\gamma$ , making it as close as possible to 1 (see paragraph “Motion estimation” for details). It can be shown that, for a given  $R$ , the optimal rule is  $S = 2\pi/(a - 1)$  (Traver and Pla 2008; Solari, Chessa, and Sabatini 2012).





The size of the RFs increases as a function of the eccentricity (the distance between the center of the RF and the fovea). The relationship between the RF size (in particular, the maximum RF size  $W_{max}$ ) and the parameters of

---

Figure 10.2 (*previous page*): **The retino-cortical mapping.** (a) Cartesian domain  $(x, y)$  with overlying log-polar pixels, i.e. the receptive fields (the circles), and (b) cortical domain  $(\xi, \eta)$ , where the squares denote the neural units. The magenta and the blue areas in (a) represent two log-polar pixels at different angular and radial positions (thus with different size  $w$  and  $h$ ) that correspond to two cortical pixels (the magenta and the blue squares in (b)) of equal size. The orange circle of RFs and the green sector of RFs in the Cartesian domain (a) map to vertical and horizontal stripes of neural units in the cortical domain (b), respectively. The red circular curve in (a) delimits the oversampling and undersampling areas, and the area inside it is the fovea. An example of image transformation from the Cartesian (d) to the cortical domain (e), and backward to the retinal domain (f). The latter is shown for completeness, though it is not used in this approach. The RFs, the yellow circles in (d), are overlying the Cartesian image. The specific choice of the parameters is:  $R = 60$ ,  $S = 93$ ,  $\rho_0 = 5$ ,  $\rho_{max} = 512$ , and  $CR = 47$ . (g) In the green box, the image transformation applied to a dead leaves stimulus image used in the experiments is shown. The specific choice of the parameters is:  $R = 110$ ,  $S = 176$ ,  $\rho_0 = 10$ ,  $\rho_{max} = 960$ , and  $CR = 47$ . The RFs and the circle delimiting the oversampling and undersampling areas are represented in black so as to not to affect the colour of the stimuli in the visualization.

the mapping is defined as follows (Solari, Chessa, and Sabatini 2012):

$$W_{max} = \rho_0 a^R (1 - a^{-1}). \quad (10.4)$$

The parameters of the log-polar mapping also influence the proportion of cortical units used to over-represent the fovea. In particular, it is possible to define the percentage of the cortical area used to represent the fovea ( $\chi$ ). This can be derived from Equation 10.4 by setting to 1 the RF size, by inverting the equation to find the corresponding  $u$  (see Equation 10.2), and by dividing for the overall size of the modeled cortex  $R$ :

$$\chi = \frac{1 - \log_a(\rho_0(a - 1))}{R}. \quad (10.5)$$

By exploiting Equations 10.4 and 10.5 we can control the growth of the size of the RFs and the over-representation of the fovea in order to reproduce data from the literature on size-to-eccentricity relationships (Freeman and Simoncelli 2011; Wurbs, Mingolla, and Yazdanbakhsh 2013; Wilkinson et al. 2016).

*Receptive field shape.* The shape of the RFs employed for the mapping affects both the quality of the transformation and its computational burden. Here, overlapping circular RFs (Bolduc and Levine 1998; Pamplona and Bernardino 2009) are employed, which are the most biologically plausible and optimally preserve image information (Chessa et al. 2011). To implement the log-polar mapping, the Cartesian plane is divided in two regions: the *fovea* and the *periphery*. The periphery is defined as the part of the plane in which the distance between the centers of two RFs on the same radius is greater than 1 pixel (undersampling). To obtain the cortical image overlapping Gaussian RFs are employed, as shown in Figure 10.2(a). The fovea (in which we have

an oversampling, i.e. the distance between two consecutive RFs is less than 1 pixel) is handled by using fixed size RFs, whereas in the periphery the size of the RFs grows. The standard deviation of the RF Gaussian profile is a third of the distance between the centers of two consecutive RFs, and the spatial support, i.e. the width of the RF, is six times the standard deviation. As a consequence of this choice, adjacent RFs overlap. A cortical pixel  $C_i$  is computed as a Gaussian weighted sum of the Cartesian pixels  $P_j$  in the  $i$ -th RF:  $C_i = \sum_j w_{ij} P_j$ , where the weights  $w_{ij}$  are the values of a normalized Gaussian centered on the  $i$ -th RF. A similar approach is used to compute the inverse log-polar mapping that produces the *retinal image*, where the space-variant effect of the log-polar mapping is observable (see Figure 10.2(c-f)). However, the model never employs the inverse log-polar mapping, since all processing is performed directly in the cortical domain.

**Motion estimation** Motion estimation, i.e. the computation of the cortical optic flow, is achieved by using a V1-MT feedforward architecture, derived from the model proposed by Simoncelli and Heeger 1998. The V1-MT feedforward architecture, for the computation of the optic flow in the Cartesian domain, has been described in Solari et al. (2015). In Solari, Chessa, and Sabatini 2012, a set of rules are presented for designing a discrete log-polar mapping that allows a direct application in the log-polar domain of the algorithms, based on spatial multi-scale and multi-orientation filtering, originally developed for the Cartesian domain without “ad hoc” modifications. The design rules are summarized as follows:

- i The aspect-ratio  $\gamma$  of the log-polar pixel has to be close to 1. This allows

the extraction of image features in the cortical domain by applying the same local operators (e.g. filtering) employed in the Cartesian domain

- ii The spatial support of the local operators has to be small on the cortical domain
- iii The mapping of a vector field of image features has to be expressed in terms of general coordinates transformation

By following these proposed rules, the distortion due to the log-polar transformation has a negligible effect on the receptive fields of the neural model cells. Thus the advantages of log-polar mapping can be exploited without the drawbacks (i.e. “ad hoc” modifications of the model) related to the distortions of the filters in the cortical domain. In particular, specific constraints are adopted on the aspect ratio of the log-polar pixel (i.e.  $\gamma = 1$ ) and on the spatial support of the filter in order to obtain undistorted RFs, since the first stage of the proposed feedforward architecture is based on spatio-temporal filtering. Thus, the optic flow can be computed from a sequence of cortical images by using the V1-MT feedforward architecture originally designed in the Cartesian domain.

Such an architecture is a three-step feedforward model: Step 1 corresponds to the V1 simple and complex cells, Step 2 corresponds to the MT pattern cells and Step 3 corresponds to a decoding stage to obtain the optical flow from the MT population response.

*Step 1:* V1 (Motion energy estimation and normalization). In the V1-layer of the model, two populations of neurons are involved in the processing of information, namely V1-direction selective simple cells and complex cells.

Simple cells are characterized by the preferred (spatial) orientation  $\theta$  of their contrast sensitivity in the spatial domain and their preferred velocity  $v^c$  in the direction orthogonal to their contrast orientation, often referred to as component speed.

The RFs of the V1 simple cells are classically modeled using band-pass filters in the spatio-temporal domain  $(x, y, t)$ . In order to achieve low computational complexity, the spatio-temporal filters  $g(x, y, \theta, f_s, f_t)$  are decomposed into separable filters in space  $h(x, y, \theta, f_s)$  and time  $p(t, f_t)$ . The spatial component of the RF is described by Gabor filters:

$$h(x, y, \theta, f_s) = Be^{\left(\frac{-(x^2+y^2)}{2\sigma^2}\right)} e^{j2\pi(f_s \cos(\theta)x + f_s \sin(\theta)y)}, \quad (10.6)$$

and the temporal component by an exponential decay function:

$$p(t, f_t) = e^{\left(-\frac{t}{\tau}\right)} e^{j2\pi(f_t t)}, \quad (10.7)$$

where  $f_s$  and  $f_t$  are the spatial and temporal peak frequencies, related to the preferred velocity by  $v^c = f_t/f_s$ , and  $\sigma$  and  $\tau$  define the spatial and temporal scales, respectively. The parameter values chosen in the current model implementation are:

- $f_s = 0.25$  cycles/pixel
- $f_t = [0, 0.10, 0.15, 0.23]$  cycles/frame (population of simple cells tuned to different preferred velocities)
- $\sigma = 2.27$  pixels
- $\tau = 2.5$  frames

The filters' spatial orientations were chosen to be:  $\theta = i\pi/8$ ,  $i = 0, 1, \dots, 7$ .

A multi-scale multi-orientation decomposition is adopted as in Freeman and Simoncelli 2011: in particular, a bank of filters of different spatial orientations is employed, and as well as a specific multi-scale approach. The multi-scale decomposition is based on a standard pyramidal approach (Burt and Adelson 1983), which can be considered as a “vertical” multi-scale, i.e. the variation of the filter size at a single location, and a “horizontal” multi-scale, where the log-polar spatial sampling produces a multi-scale for different spatial locations, i.e. the variation of the filter size across the visual field (Bonmassar and Schwartz 1997).

In order to process the cortically-transformed images, it is necessary to characterize the filters, defined in the Cartesian domain, with respect to the cortical domain, i.e. to map the filters into the cortical domain, thus obtaining  $g(x(\xi, \eta), y(\xi, \eta), \theta, f_s, f_t)$ . As a consequence of the non-linearity of the log-polar mapping, the mapped filters are distorted, and in Solari, Chessa, and Sabatini (2012) it has been shown that under specific conditions such distortions can be kept to a minimum. This happens when the spatial support of the RFs is sufficiently small, and the aspect ratio  $\gamma$  of the log-polar pixel is equal to 1, see Figure 10.3 (top). Under these assumptions, it is possible to directly work in the cortical domain, by considering spatio-temporal filters sampled in log-polar coordinates  $g(\xi, \eta, \theta, f_s, f_t)$ , see Figure 10.3 (bottom).

Given the response of the simple cells layer, the V1 complex cells are described as a combination of the quadrature pair of simple cells by using the motion energy formulation (Adelson and Bergen 1985), followed by a

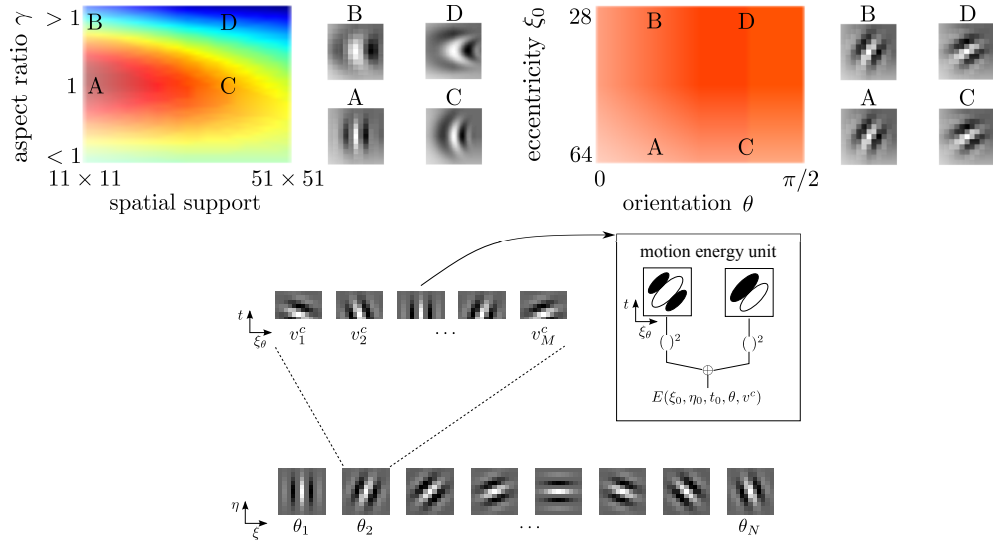


Figure 10.3: **Filtering in the cortical domain.** (top) Variations of the energy ratio between mapped  $g(x(\xi, \eta), y(\xi, \eta), \theta, f_s, f_t)$  and matched  $g(\xi, \eta, \theta, f_s, f_t)$  filters, with respect to (left) the spatial support of the filters and the aspect ratio  $\gamma$  of the log-polar pixel; (right) the orientation  $\theta$  of the filters and the eccentricity in the cortical plane  $\xi$ , when  $\gamma = 1$  and spatial support  $11 \times 11$  pixels. The profiles of the mapped filters for particular choices of parameters are marked by letters A - D. Hot colors are high energy ratios (i.e. the ratio between mapped and matched filter is close to 1, thus the distortions of the mapped filter are minimal), cold colors mean low energy ratios. With aspect ratio  $\gamma = 1$  and spatial support  $11 \times 11$  pixels, distortions are minimal for every orientation  $\theta$  and eccentricity  $\xi_0$ . (bottom) Spatio-temporal filters sampled in log-polar coordinates ( $g(\xi, \eta, \theta, f_s, f_t)$ ), tiling  $N$  orientations  $\theta$ . For each orientation  $\theta$ ,  $M$  tuning velocities are considered. Top row shows Gabor filters in the  $(\xi_\theta, t)$  plane for a given  $\theta_i$  (with  $t > 0$ , since the temporal filter are causal). The inset on the right describes a motion energy unit.



divisive normalization (Heeger 1992), denoted as  $E^{V1}(\xi, \eta, t, \theta, v^c)$ . A key property of V1 cells is their tuning to the spatial orientation and velocity of a stimulus, which arises from spatio-temporal frequency selectivity for motion in a direction perpendicular to the contrast of the underlying pattern (Adelson and Movshon 1982).

*Step 2:* MT pattern cells response. MT neurons exhibit velocity tuning irrespective of the contrast orientation. This is believed to be achieved by pooling afferent responses from V1 layers, in both spatial and orientation domains, followed by a non-linearity. In particular, in the proposed model we perform the following processing:

- i the output of the V1 afferent cells is spatially pooled through a Gaussian kernel
- ii the previous output is pooled by MT linear weights, which give rise to the MT tuning to speed direction  $d$ ; these weights are defined as  $\cos(d - \theta)$  where  $d \in [0, 2\pi]$
- iii the output of the MT orientation pooling is then fed into an exponential function which describes the static non-linearity.

The responses of an MT pattern cell tuned to the speed  $v^c$  and to direction of speed  $d$  is denoted as  $E^{MT}(\xi, \eta, t, d, v^c)$ .

*Step 3:* Decoding. In this step, optic flow is estimated by decoding the population responses of the MT neurons. Indeed, a unique velocity vector cannot be recovered by the activity of a single velocity tuned MT neuron as multiple scenarios could evoke the same activity, but a unique vector can be recovered based on the activity of a population. Here, a linear combination

approach is employed to decode the MT population response as described previously (Pouget et al. 1998; Rad and Paninski 2011).

The MT representation of the speeds  $v^c$  and the speed directions  $d$  are linearly decoded: first it is possible to decode the MT responses  $E^{MT}(\xi, \eta, t, d, v^c)$  along each speed direction  $d$  to compute the speed, then to apply the intersection of constraints on such estimated velocities.

The estimates  $v^d$  of the speed along each speed direction  $d$  can be obtained as follows (i.e. linear combination):

$$v^d(\xi, \eta, t, d) = \sum_{v_i^c=v_1^c}^{v_M^c} v_i^c E^{MT}(\xi, \eta, t, d, v_i^c) \quad (10.8)$$

Then, the estimate of the full cortical velocity is:

$$\begin{aligned} v_\xi &= \frac{2}{Q} \sum_{d_i=d_1}^{d_Q} v^d(\xi, \eta, t, d_i) \cos d_i \\ v_\eta &= \frac{2}{Q} \sum_{d_i=d_1}^{d_Q} v^d(\xi, \eta, t, d_i) \sin d_i. \end{aligned} \quad (10.9)$$

It is worth noting that an expansion in the Cartesian domain is represented by a constant flow along the  $\xi$  axis in the cortical domain, but if the FRM is shifted with respect to the center of the visual field (i.e. we have an expansion with a constant translation), the corresponding cortical flow has non-linear components, as shown in Figure 10.4. Moreover, the cortical flow generated by a shift of the FRM towards the right is notably different from the cortical flow corresponding to a shift of the FRM towards the left.

**Motion interpretation** Given the sensitivity of MSTd neurons for Elementary Flow Components (EFCs) alone or their combination with translation components (Orban et al. 1992), here we may consider a population of cells, which model the tuning to EFCs (see Figure 10.5).

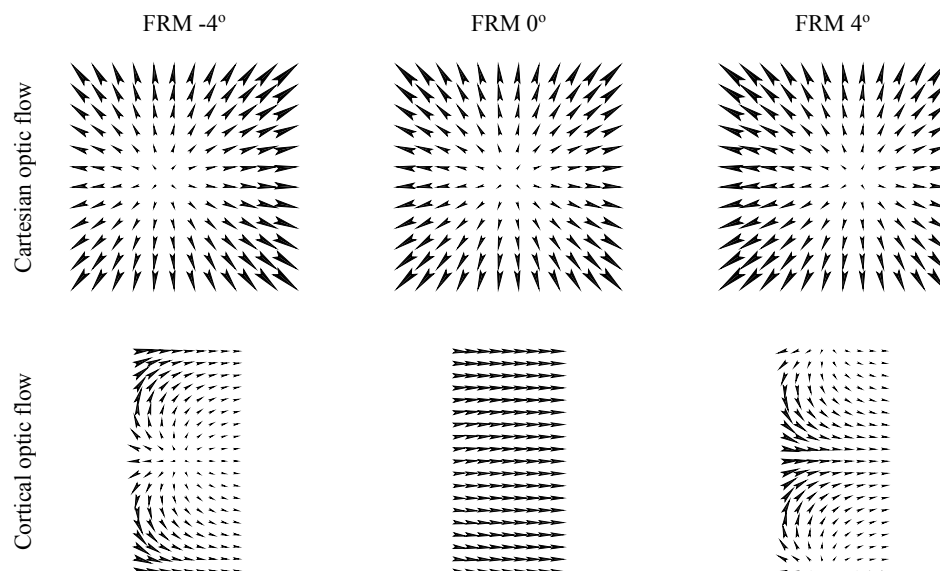


Figure 10.4: **Optic flow in the cortical domain.** (top) Optic flows representing expansion (or divergence) in the Cartesian domain, with different shifts of the FRM, and (bottom) corresponding cortical optic flows. Small ( $\pm 4^\circ$ ) variations of the FRM location produce high non-linearities in the cortical flows.

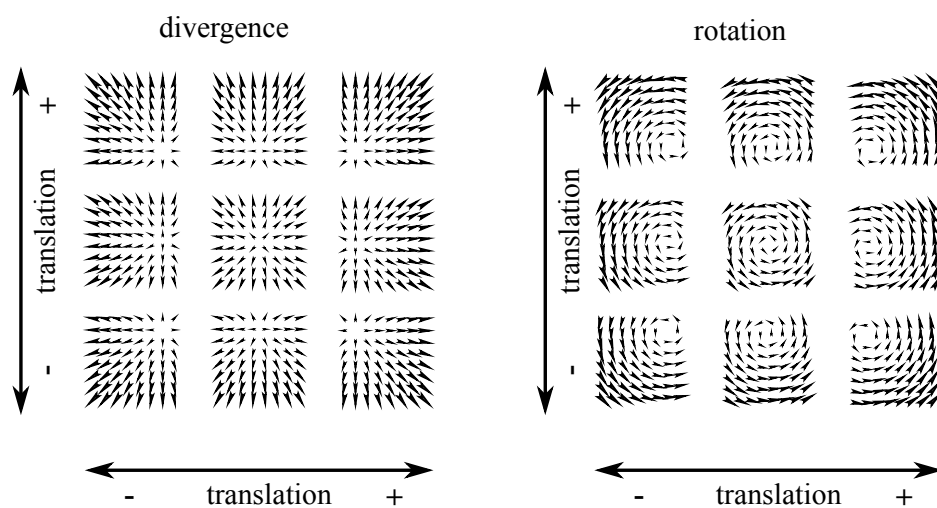


Figure 10.5: **Deformation subspaces.** Two deformation subspaces (Chessa, Solari, and Sabatini 2013), representing an expansion (left) and a rotation (right), obtained from the combination of deformation gradients and translation components.

In particular, by considering the Cartesian (i.e. retinal) space, we have four classes of deformation gradients: one stretching and one shearing, for each cardinal direction; and two translation components. The deformation and translation templates can be combined to obtain deformation subspaces representing EFCs such as expansion, shear and rotation. It is worth noting that the expansion and rotation subspace represented in Figure 10.5 are slices that can be obtained from the *Spiral space model* presented in Xu, Wallisch, and Bradley (2014).

In the Cartesian domain, the EFCs can be described in terms of affine description (for details see Chessa, Solari, and Sabatini 2013), and Solari, Chessa, and Sabatini (2014) show that the affine description in the Cartesian domain can be recovered from the cortical affine description. Moreover, in that paper the relationships among the first order description of the cortical flow and the estimation of the 3D rigid body motions have been derived. Thus, it is possible to recover a description of the EFCs by working directly on the cortical optic flow.

The cortical motion field can be described as linear deformations by a first-order Taylor decomposition, around each cortical image point:  $\mathbf{v} = \bar{\mathbf{v}} + \bar{\mathbf{T}} \cdot [\xi, \eta]^T$ , where  $\bar{\mathbf{T}}$  is the tensor composed of the partial derivatives of the cortical motion field. By describing the tensor through its dyadic components, we obtain:

$$\mathbf{v} = \alpha^\xi \bar{v}_\xi + \alpha^\eta \bar{v}_\eta + \mathbf{d}_\xi^\xi \left. \frac{\partial v_\xi}{\partial \xi} \right|_{(\xi_0, \eta_0)} + \mathbf{d}_\eta^\xi \left. \frac{\partial v_\xi}{\partial \eta} \right|_{(\xi_0, \eta_0)} + \mathbf{d}_\xi^\eta \left. \frac{\partial v_\eta}{\partial \xi} \right|_{(\xi_0, \eta_0)} + \mathbf{d}_\eta^\eta \left. \frac{\partial v_\eta}{\partial \eta} \right|_{(\xi_0, \eta_0)}, \quad (10.10)$$

where  $\alpha^\xi : (\xi, \eta) \mapsto (1, 0)$ ,  $\alpha^\eta : (\xi, \eta) \mapsto (0, 1)$  are pure translations and  $\mathbf{d}_\xi^\xi : (\xi, \eta) \mapsto (\xi, 0)$ ,  $\mathbf{d}_\eta^\xi : (\xi, \eta) \mapsto (\eta, 0)$ ,  $\mathbf{d}_\xi^\eta : (\xi, \eta) \mapsto (0, \xi)$ ,  $\mathbf{d}_\eta^\eta : (\xi, \eta) \mapsto (0, \eta)$

$(\xi, \eta) \mapsto (0, \eta)$  represent the cardinal deformations (gradients), i.e the basis of the linear deformation space, in the cortical domain. Figure 10.6 shows the cardinal deformations: the stretching and shearing gradients along each directions, and the two translation components.

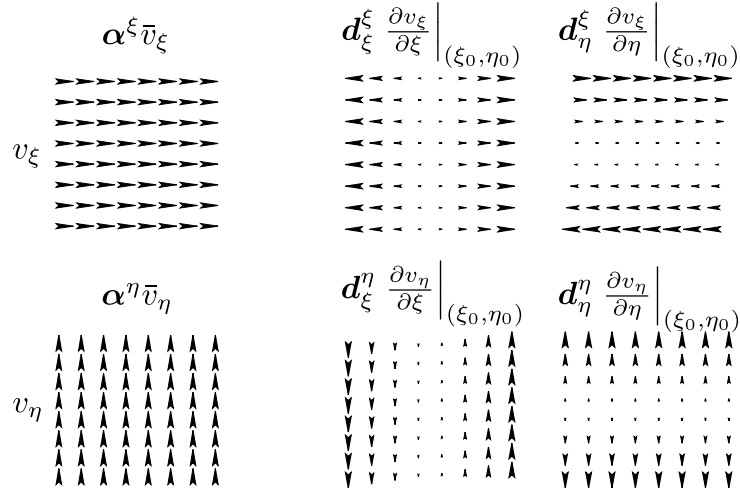


Figure 10.6: **Local cardinal deformations of the cortical optic flow.** The local cardinal deformations of the cortical optic flow (see Equation 10.10), basis of its first-order approximation around a cortical point  $(\xi_0, \eta_0)$ . Such basis are the templates used to perform the template matching on the cortical optic flow, to compute the affine description  $[\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_6]$  of a local patch (see Equation 10.11).

The sensitivity to such deformations can be modeled as a population of cells, whose response is computed through an adaptive template matching on the cortical optic flow. From the responses of such a population we may compute the first order (affine) description (Koenderink 1986) of the cortical

optic flow:

$$\begin{bmatrix} v_\xi \\ v_\eta \end{bmatrix} = \begin{bmatrix} \tilde{c}_1 \\ \tilde{c}_4 \end{bmatrix} + \begin{bmatrix} \tilde{c}_2 & \tilde{c}_3 \\ \tilde{c}_5 & \tilde{c}_6 \end{bmatrix} \cdot \begin{bmatrix} \xi \\ \eta \end{bmatrix}, \quad (10.11)$$

where  $\tilde{c}_i$  are constants and  $v_\xi$  and  $v_\eta$  are the components of the cortical optic flow. The parameter vector  $[\tilde{c}_1, \tilde{c}_2, \dots, \tilde{c}_6]$  describes a specific configuration of cortical optic flow in a local patch.

Motion interpretation (Xiao, Marcar, and Raiguel 1997) (i.e. the estimation of the 3D orientation of the surfaces, of the time-to-collision and of the FRM) is performed with respect to the Cartesian coordinates. Indeed, the “interaction” with the real-world occurs in Cartesian coordinates, thus we need the relationships among the cortical descriptors and the Cartesian ones in the area of interest in order to perform the motion interpretation by using the cortical first order descriptors, as it happens in the visual cortex.

In Solari, Chessa, and Sabatini (2014), the mathematical formulations are derived that allow us to recover time-to-contact and the surface orientation information from the cortical optic flow. Similarly the mathematical formulation that can be employed to estimate the FRM from the cortical optic flow may be derived (Chessa, Maiello, Bex, et al. 2016). By following the *Spiral space model*, the location of the FRM is computed by detecting the area where the optic flow can be locally described by a divergence (i.e. expansion) component and a “null” translational (i.e. laminar motion) component. In particular, by considering the cortical first-order description, the FRM is computed as follows. To detect a null translational component we have to locate the retinal region where the Cartesian affine coefficients  $c_1$  and  $c_4$  (see Chessa, Maiello, Bex, et al. 2016) are null. We can reformulate this

rule in the cortical domain by using the relationships described in Chessa, Maiello, Bex, et al. (2016), which yields:

$$\tilde{c}_2^2 + \tilde{c}_5^2 < \frac{T_1}{\rho_0^2 a^{2\xi}}, \quad (10.12)$$

where  $T_1$  denotes a threshold value. It is worth noting that the right side of Equation 10.12 is not a constant, i.e. it decreases as a function of  $\xi$ .

All cortical regions which satisfy Equation 10.12 will correspond to retinal locations with negligible translational optic flow. Next, we must locate the retinal locations where the optic flow is expanding or contracting, and can thus be described as a non null divergence, i.e. the Cartesian affine coefficients must obey the rule  $|c_2 + c_6| > 0$ . By using the relationships described in Chessa, Maiello, Bex, et al. (2016) this rule can be reformulated with affine cortical coefficients as:

$$\tilde{c}_1^2 + \tilde{c}_2^2 + \tilde{c}_6^2 > \frac{T_2}{\ln(a^2) \ln(a) + 1} = T_3, \quad (10.13)$$

where  $T_2$  ( $T_3$ ) denotes a threshold value, and  $a > 1$ . The use of thresholds, instead of searching for null translational and non-null divergent flow components, allows us to handle the variability present in the visual stimulus in real-world situations (Chessa, Maiello, Bex, et al. 2016). In the simulations carried out here:  $T_1 = 0.8$  and  $T_2 = 0.6$ .

**Model Implementation** A preliminary version of the proposed model has been implemented in C++, and is capable of real-time performance (Solari, Chessa, and Sabatini 2014). For the experiments described in this chapter I employed a MATLAB implementation of the proposed model. Nine frames



are fed to the model each trial. Because the model does not process color information in any way, RGB image sequences were fed to the model in gray-scale. The *Retino-cortical mapping* module uses as inputs images  $1920 \times 1080$  pixels and produces cortical images  $214 \times 150$  pixels, thus achieving a compression ratio of  $CR = 64.6$ . Then the cortical images are processed by the *Motion estimation* module: the spatio-temporal V1 RFs' size is  $11 \times 11 \times 5$ , i.e. the spatial support is  $11 \times 11$  pixels and the temporal support is 5 frames, near the fovea, the considered spatial orientations are 16, the number of spatial scales is 6, the V1 tuning velocities are 9 between -0.9 and 0.9 pixel/frame. Finally, the *Motion interpretation* (MST) module processes the cortical optic flows by using larger RFs ( $48 \times 48$  pixels near the fovea).

**FRM Estimation as a Function of Eccentricity** The model described thus far implements known aspects of the first stages of the dorsal cortical pathway for motion processing, and is potentially able to estimate the FRM in image sequences containing motion. I thus tested the model's performance at FRM Estimation, and compared the model's performance to the performance of human observers. I was interested in whether the proposed model would capture the changes in human performance at motion estimation across the visual field. If the model were to show similar patterns of performance to human observers, we could take that as good evidence that the model implementation is consistent with the neural computations performed by the visual system regarding motion estimation, by exploiting a logic similar to that employed by Freeman and Simoncelli (2011). If the model were unable to perform the motion estimation task, or if the pattern

of performance differed between model and human observers, we could conclude that the implemented aspects of known neural motion processing are insufficient to fully describe how the human visual system extracts motion information from the cortical visual representation of the world.

Broadly, it is well established that motion sensitivity co-varies with spatial resolution in the peripheral visual field (McKee and Nakayama 1984; Wright and Johnston 1983). More specifically, Bex and Falkenberg (2006) have shown that the accuracy with which human observers can estimate the FRM in random dot stimuli decreases as a function of retinal eccentricity. I proceeded to replicate the findings from Bex and Falkenberg (2006) in human observers with more naturalistic dead leaves stimuli (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001). I then tested the model on the same task and stimuli, and compared the model’s performance to that of the human observers.

**Equivalent Noise Analysis of FRM Estimation as a Function of Eccentricity** Two main factors may affect the accuracy and precision with which humans estimate the FRM in complex, noisy motion stimuli. Loss of resolution in the peripheral visual field may add noise to local direction estimates on which global optic flow computations are performed. Additionally, receptive field size differences throughout the visual field may affect the efficiency with which motion information present within a stimulus is integrated. Resolution loss and receptive field size changes throughout the visual field are built into the proposed model. Thus, I ask whether these features of the model produce the same effects as they do in human observers.

To tease apart the contribution of internal noise and integration efficiency to the precision of FRM estimation, I employ the Equivalent Noise (EN) paradigm (Barlow 1956; D. Pelli 1990) as employed by Bex and Falkenberg (2006). The EN paradigm is based on the assumption that an observer’s performance is limited by additive internal noise as well as by how efficiently the observer samples the information available from the stimulus. By assuming that the variance in the stimulus and variance in the visual system are additive, thresholds for FRM estimation can be expressed as:

$$\sigma_{FRM} = \sqrt{\frac{\sigma_{int}^2 + \sigma_{ext}^2}{n}} \quad (10.14)$$

where  $\sigma_{FRM}$  is the FRM discrimination threshold,  $\sigma_{int}$  is the internal noise of the system,  $\sigma_{ext}$  is the external noise contained in the stimulus, and  $n$  is the sampling efficiency, which relates to how well the system is able to integrate the information contained within the stimulus.

I perform EN analysis on both human observers and on the model, and compare the results. If the pattern of results in humans and in the model are compatible, this would suggest that the implemented features of the model regarding resolution loss and receptive field size are having similar effects on the performances of humans and model.

## 10.2 Methods

### Experiment 1: FRM Estimation in Human Observers

**Participants** Two expert (GM and PB) and one naive observer (WH) were recruited to participate in Experiment 1. All three subjects were male and

were 27, 49 and 31 years of age respectively with normal or corrected vision in the test eye.

**Stimuli** An example frame of generated stimuli can be seen in Figure 10.7. Stimuli were expanding or contracting dead leaves patterns (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001) generated using Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997) running on MATLAB (Mathworks). Dead leaves were constructed from a set of 2000 limited-lifetime, 2D Gaussian ellipses, each drawn with a random orientation, aspect ratio, and RGB color, randomly placed within a circular region with diameter equal to half monitor length. FRM sensitivity is relatively invariant to element lifetime (Warren Jr et al. 1991), thus to prevent observers from tracking single elements over time, the lifetime of the ellipses was fixed at five movie frames. The limited lifetime also prevented large density changes that may otherwise occur as ellipses cluster in the center of the image with contracting motion. On the first movie frame the ellipses were assigned a random starting location and a random age between one and five movie frames which ensured that all elements would not expire simultaneously. The location of each ellipse was updated at every movie frame. The direction  $dot_{Dir}$  of motion of each ellipse was computed as:

$$dot_{Dir} = \text{atan2}(y_{Dot} - y_{FRM}, x_{Dot} - x_{FRM}) + Motion_{Dir}, \quad (10.15)$$

where  $\text{atan2}(y, x)$  is the four quadrant inverse tangent function,  $(x_{Dot}, y_{Dot})$  and  $(x_{FRM}, y_{FRM})$  are the  $(x, y)$  coordinates of the ellipse and FRM respectively, and  $Motion_{Dir}$  is 0 radians for expanding and  $\pi$  radians for contracting motion.

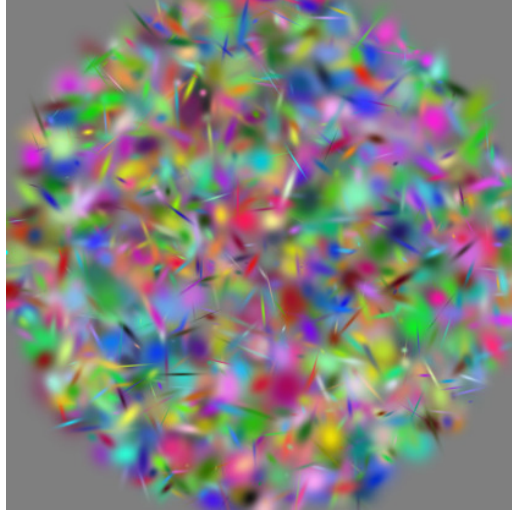


Figure 10.7: **Dead leaf stimuli.** Example single frame from the stimuli, which were expanding or contracting fields of dead leaves.

The updated  $(x, y)$  position of each ellipse was then computed as:

$$\begin{aligned} x_{Dot} &= x_{Dot} + \cos(dot_{Dir}) * dot_{Dist} * dot_{Speed} \\ y_{Dot} &= y_{Dot} + \sin(dot_{Dir}) * dot_{Dist} * dot_{Speed} \end{aligned} \quad (10.16)$$

where  $dot_{Dist}$  is the distance of the ellipse from the FRM and  $dot_{Speed}$  is the speed of an ellipse one pixel away from the FRM, which was set to be 0.1 pixel/frame. These computations generated expanding or contracting motion with a realistic speed gradient. Elements that fell outside the circular stimulus region or whose lifetime exceeded five movie frames were randomly repositioned within the stimulus and assigned an age of zero.

**Apparatus** FRM stimuli were presented on a Dell P2815Q monitor with resolution 1920 by 1080 pixels run at 60 Hz from an AMD Radeon HD 7000 graphics processing unit. Subjects were positioned 50 cm from the monitor,

which subtended 64x38 degrees of visual angle. Observers were positioned on a chin and forehead rest, with their non-dominant eye occluded, so as to view the stimuli monocularly and reduce motion and depth cue conflicts. Fixation compliance was monitored using a Tobii EyeX eyetracker, a low-cost eyetracker validated for research purposes in Chapter 11 and in Gibaldi et al. (2016).

**Design** To replicate the findings from Bex and Falkenberg (2006) I measured the accuracy of motion integration by asking human observers to position the mouse cursor at the perceived location of the FRM presented at different locations into the peripheral visual field. Each observer completed 150 trials. In random order, in one third of the trials the FRM appeared at 4 degrees, in one third it appeared at 8 degrees, and in one third it appeared at 12 degrees. To minimize the buildup of motion aftereffects, the direction of motion (expansion or contraction) was randomly assigned each trial. In post-processing, a total of 19% trials were removed in which the eyetracker data signaled the observers had not maintained steady fixation within 3 degrees of the central fixation target.

**Procedure** Observers were instructed to fixate a central fixation dot, while stimuli were presented to their nasal visual field (in order to avoid the physiological blind spot). The stimulus, a circular dead leaves patch generated as described above, subtended 35 degrees of visual angle and was centered 17.5 degrees to the right of fixation. At the start of each trial, the observers were presented with a fixation dot and were shown their own gaze point (estimated through the use of the eyetracker) to remind them to maintain steady central

fixation. When ready, observers initiated a trial by mouse click. Within the circular stimulus, the FRM was randomly selected along 30 degree arcs at either 4, 8 or 12 degrees from central fixation (see Figure 10.9a, black asterisks indicate FRM test locations). Motion stimuli were presented for 120 frames (2 seconds), during which time the mouse cursor was not presented. Once the stimulus had been extinguished the mouse cursor was presented on screen at a random position on a uniform gray background. Observers were free to move their eyes and their task was to move the tip of the cursor to the perceived FRM and press the mouse button.

## **Experiment 2: FRM Estimation in the Proposed Model**

To compare the performance of the human observers to the performance of the model, the same expanding and contracting dead leaves stimuli were fed to the model. The output of the computations performed by the model on each image sequence was the estimated FRM for that sequence. I ran 150 image sequences through the model, and recorded its estimated FRM for each sequence.

## **Experiment 3: Equivalent Noise Analysis in Human Observers**

**Participants** Six human observers were recruited to participate in Experiment 3 (two expert observers and 4 naive participants, mean age  $34 \pm 9$  years, 5 males and 1 female). All observers had normal or corrected to normal vision in the test eye.

**Stimuli** Stimuli were the same expanding and contracting dead leave patches employed in the previous set of experiments. External noise was added to the stimuli as follows. Instead of selecting a single FRM position for all dead leaf elements, the  $(x, y)$  positions of the FRM of each element were selected from normal distributions with means equal to  $(x_{FRM}, y_{FRM})$  and standard deviation equal to  $\sigma_{ext}$ , i.e. the external noise.

**Apparatus** FRM stimuli were presented on an ASUS VG278HE monitor with resolution 1920 by 1080 pixels run at 60 Hz from an NVidia Quadro 580 graphics processing unit. Subjects were positioned 55cm from the monitor, which subtended 57x34 degrees of visual angle. Observers were positioned on a chin and forehead rest, with their non-dominant eye occluded, so as to view the stimuli monocularly and reduce motion and depth cue conflicts. Observers were instructed to fixate a central fixation dot while the stimuli, circular dead leaves patches generated as described above, subtended 34 degrees of visual angle and were centered at fixation. As previously, fixation compliance was monitored using a Tobii EyeX eyetracker (Gibaldi et al. 2016).

**Design** The full EN function is typically estimated by measuring observers' thresholds (in this case the FRM discrimination threshold  $\sigma_{FRM}$ ) at varying amounts of external noise ( $\sigma_{ext}$ ), and the observed thresholds are then fit to Equation 10.14, thus obtaining estimates of the internal noise ( $\sigma_{int}$ ) and sampling efficiency ( $n$ ) parameters. FRM discrimination thresholds at each tested eccentricity were thus measured at five fixed levels of external noise: 0.25, 0.5, 1, 2 and 4 degrees. The thresholds were measured via 15



randomly interleaved staircases (Wetherill and Levitt 1965). The raw data from a minimum of 50 trials from each staircase were combined and fitted with a cumulative normal function by weighted least-squares regression (in which the data are weighted by their binomial standard deviation). FRM discrimination thresholds were estimated from the 80% correct point of the psychometric function. For each tested eccentricity these thresholds were fit via nonlinear least squares regression to the EN function presented in Equation 10.14.

**Procedure** The sequence of events from a single trial is shown in Figure 10.8. At the start of each trial, the observers were presented with a fixation dot and after 250 ms were also shown a cross at either the fovea, at 4 or at 8 degrees eccentricity, to cue where the stimulus would appear in the visual field. The cue was extinguished after 500 ms, and motion stimuli were then presented for 9 frames (150 ms). The FRM were located at the fovea or at 4 or 8 degrees eccentricity, shifted into one of four quadrants of the image patch by an amount that was under the control of a 3 down, 1 up staircase (Wetherill and Levitt 1965) that adjusted the shift to a level that produced 79% correct trials. The testing procedure's goal was to estimate the minimum FRM shift necessary for observers to identify in which of the four quadrants the FRM had been presented. Once the stimulus had been extinguished observers were required to indicate, via mouse click, in which of the four image quadrants centered at the eccentric testing location they had perceived the FRM. Observers were given unlimited time to respond. The following trial commenced as soon as observers provided a response.

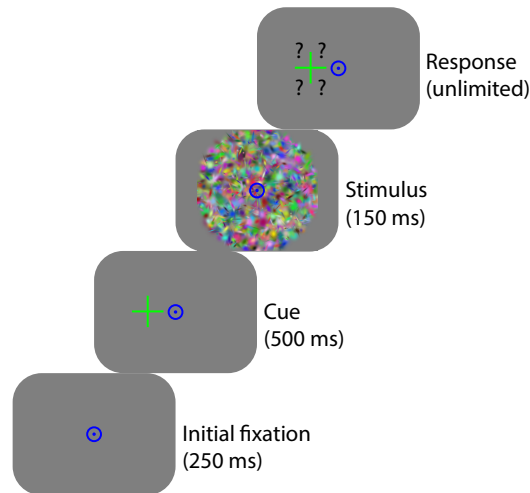


Figure 10.8: **Schematic of a single trial from equivalent noise experiment.** Observers were required to fixate a central fixation target (blue). While maintaining steady fixation, observers were shown a green cross at either the fovea, at 4 or at 8 degrees eccentricity, to cue approximately where the FRM would appear in the visual field. Fixation compliance was ensured with an eye tracker. The stimulus, a full field of expanding or contracting dead leaves, then appeared on screen for 150 ms, an interval that is too brief for a change in fixation. Following the stimulus presentation, observers were required to indicate, via mouse click, in which of the four image quadrants centered at the eccentric testing location they had perceived the FRM (question marks are for illustration only and were not present in the experiment).

## **Experiment 4: Equivalent Noise Analysis in the Proposed Model**

In order to compare the performance of the human observers to the performance of the model, the same expanding and contracting dead leaves stimuli were fed to the model. The stimuli, testing procedure and data processing employed on the model were the same as those employed with human observers in Experiment 3. Instead of a mouse click, the model output was an estimate of the FRM location within its visual field. A single trial from the model was scored as correct if the estimate fell within the correct stimulus quadrant. The trial was scored as incorrect otherwise. If for a single trial the model was unable to return an estimate of the FRM, the trial was repeated with a new image sequence.

EN functions at each eccentricity were estimated from seven separate model parameter configurations. In each of the seven experimental runs, the model parameters were pseudo-randomly chosen from within a sensible range. The parameter configurations employed in each run are reported in Table 10.1.

## **10.3 Results**

### **Experiments 1 and 2: FRM Estimation**

Figure 10.9 shows the results of the FRM Estimation Experiment 1 (from the three human observers) and Experiment 2 (from the model). Human observer data were discarded when the eyetracker data reported incorrect fixations,

Table 10.1: **Model parameters for Equivalent Noise experiment.** For the seven experimental runs (I-VII) the the number of rings  $R$  and the blind spot radius  $\rho_0$  were varied, thus obtaining the corresponding percentage of cortex used to over-represent the fovea  $\chi$  (see Equation 10.5), the compression ratio  $CR$  (see Equation 10.3), and the maximum RF size  $W_{max}$  (see Equation 10.4).

Run No	$R$	$\rho_0$	$\chi$	$CR$	$W_{max}$
I	130	9	33	35	14
II	160	7	40	24	12
III	110	10	28	47	16
IV	110	3	40	63	21
V	160	3	47	29	15
VI	130	3	43	45	18
VII	105	9	27	54	18

or when the monitor refresh rate fell below 50 Hz, which would occasionally occur due to the high computational costs of generating the dead leaves stimuli in real time. This resulted in 19% of discarded trials from observers. Data from the model were also discarded when the thresholds in Equations 10.12 and 10.13 were not reached and thus the model was unable to output a FRM estimate. This resulted in 21% of discarded trials from the model.

Figure 10.9 presents the error vectors between the actual FRM test locations and the perceived FRM locations in the three human observers (a) and the estimated FRM locations from the model (b). Qualitatively we note that both the model and the human observers can perform the task, although with some degree of error. Furthermore, the spread of the data across human observers and model is qualitatively similar. Figure 10.9(c) shows the average perceived location of the FRM at each eccentricity for the three observers and the model. All three observers exhibited a consistent foveocentric bias at all three test eccentricities. Similar biases have been previously observed in FRM estimation tasks (Johnston, White, and Cumming 1973; Warren Jr and Saunders 1995) but were not observed by Bex and Falkenberg (2006). Conversely to the human observers in some studies, the model does not exhibit a consistent bias.

Figure 10.9(d) shows average precision errors, computed as the absolute distance between the target and perceived FRM, as a function of eccentricity for the three observers and for the model. Errors in human observers ranged between 1-3 degrees and increased as a function of eccentricity, in good agreement with the literature (Warren Jr, Morris, and Kalish 1988; Warren Jr and Saunders 1995) and especially with Bex and Falkenberg (2006), in spite

of differences between the stimuli employed in the different studies. The model's absolute error in FRM estimation as a function of retinal eccentricity is similar, both in magnitude and trend, to the data from human observers. Given these results, we can conclude that the theoretically based neural computations implemented in the proposed model are consistent with the computations performed by the human visual system regarding complex motion estimation.

### Experiments 3 and 4: Equivalent Noise Analysis

Figure 10.10 summarizes the results of the EN analysis performed in six human observers. Figure 10.10a shows FRM threshold offsets as a function of external noise for all six observers. Red, green and blue curves passing through the data are the average fitted EN curves at each eccentricity. The average estimated EN parameters are plotted as a function of eccentricity in Figures 10.10b (Internal Noise) and 10.10c (Sampling Efficiency). Error bars are 95% bootstrapped confidence intervals of the mean parameter estimates.

Figure 10.11 presents the results of the EN analysis performed on seven separate model parameter configurations overlain onto the observed range of human performance. Figure 10.11a shows FRM threshold offsets for all seven model parameter configurations and fitted EN curves as a function of external noise, averaged across the seven model parameter configurations at each eccentricity. The average estimated EN parameters for the model are plotted as a function of eccentricity in Figures 10.11b (Internal Noise) and 10.11c (Sampling Efficiency). Average EN parameter estimates bounded by 95% bootstrapped confidence intervals are overlain onto gray shaded 95%

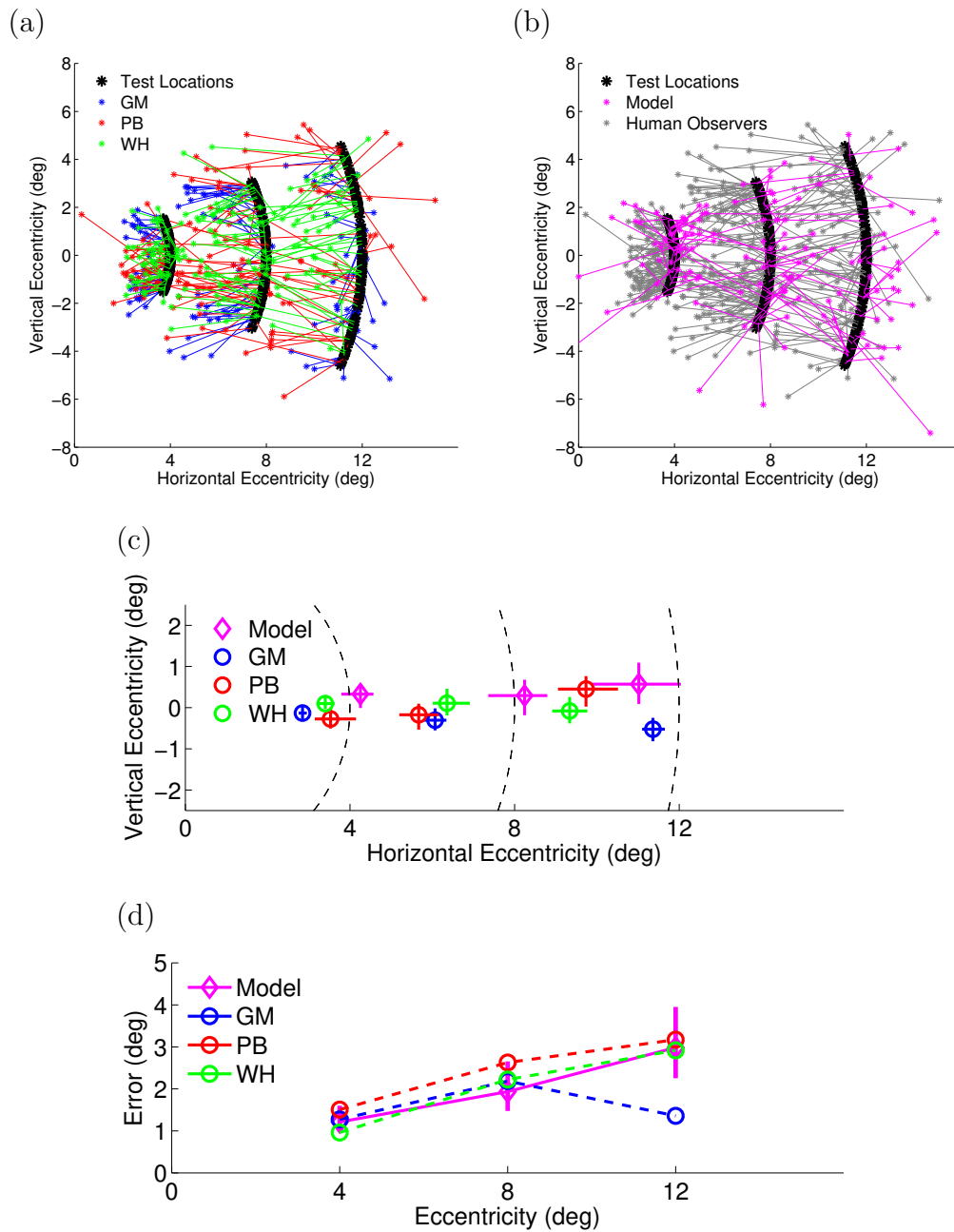


Figure 10.9

---

Figure 10.9 (*previous page*): **FRM estimation in human observers and proposed model.** (a) Error vectors between true and perceived FRM location for every trial throughout the visual field of human observers. Eccentricity is plotted in degrees from central fixation. Black asterisks are true FRM test locations. Blue, red and green asterisks are perceived FRM locations from the three human observers. Each colored asterisk is connected to its true FRM location by a straight line. (b) Error vectors between true and estimated FRM location for every trial throughout the visual field of the model. Magenta asterisks are estimated FRM locations from the model superimposed onto non-individualized gray asterisks corresponding to the human observer error vectors from (a). (c) Perceived FRM location at each eccentricity for three observers and the model. Each data point shows the mean perceived FRM relative to the true FRM at each tested eccentricity. Error bars are 95% bootstrapped confidence intervals. Color coding is as in previous figure. (d) Mean absolute error between the estimated/perceived and the actual FRM as a function of eccentricity. Dotted blue, red and green lines are mean error for the three subjects. Filled magenta line with error bars is mean error for the model. Error bars are 95% bootstrapped confidence intervals.



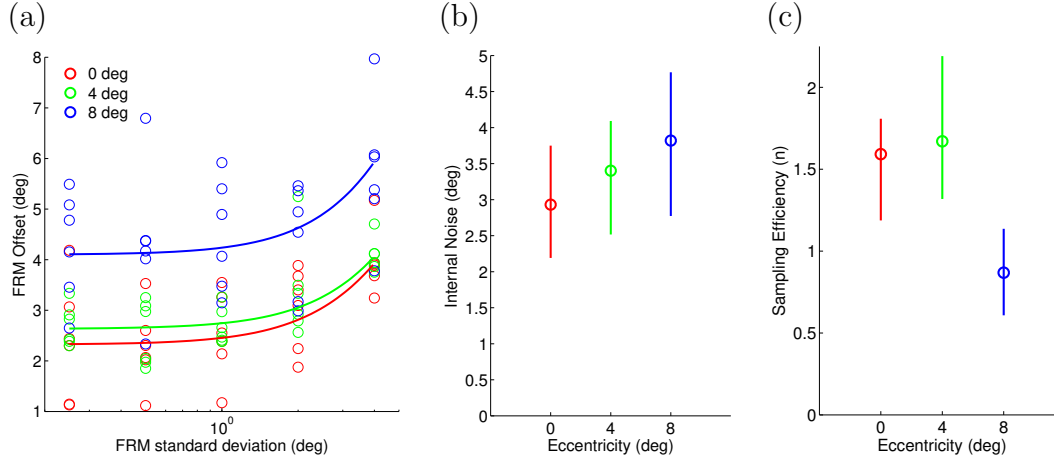


Figure 10.10: **Equivalent Noise Analysis in human observers.** (a) Equivalent noise functions at fovea (red), 4 degrees (green) and 8 degrees (blue) eccentricity. The data show FRM discrimination thresholds as a function of positional noise applied to each element within the stimulus. Individual data points are discrimination thresholds for all six observers. Curves are best fitting equivalent noise functions to the averaged data across the six observers. Note that  $x$ -axis is log scaled. (b) Internal Noise and (c) Sampling Efficiency parameters of the estimated equivalent noise functions are plotted as a function of eccentricity. Data are the average across six observers. Error bars represent 95% bootstrapped confidence intervals

confidence regions for the human EN parameter estimates.

In both human observers (Figure 10.10a) and model (Figure 10.11a), thresholds vary lawfully as a function of external noise. FRM location identification performance is better at low levels of external noise than at high. Furthermore, performance worsens with increasing eccentricity. The model's performance is better than that of human observers for low levels of external noise, but worsens at a faster rate with increasing levels of external noise.

To statistically evaluate how Internal Noise and Sampling Efficiency vary throughout the visual field of human and model observers, EN parameter estimates were analyzed via a 2 (observer type) x 3 (eccentricity) ANOVA.

The ANOVA result for Internal Noise showed a significant main effect of observer type,  $F(1,33)=61.12$ ,  $p=10^{-8}$ , no significant main effect of eccentricity,  $F(2,33)=1.34$ ,  $p=0.28$ , and no significant interaction between observer type and eccentricity  $F(2,33)=0.39$ ,  $p=0.68$ . As can be seen in Figures 10.10b and 10.11b, there is a trend for Internal Noise in both human and model observers to increase with eccentricity, however the increase is not statistically significant. The model has significantly less overall Internal Noise than human observers.

The ANOVA result for Sampling Efficiency showed a significant main effect of observer type,  $F(1,33)=20.45$ ,  $p=10^{-4}$ , a significant main effect of eccentricity,  $F(2,33)=17.94$ ,  $p=10^{-5}$ , and a significant interaction between observer type and eccentricity  $F(2,33)=4.39$ ,  $p=0.02$ . Sampling Efficiency in both human (Figures 10.10c) and model observers (Figures 10.11c) decreases with eccentricity. However, the model has significantly lower overall Sampling Efficiency than human observers. Furthermore, Sampling Efficiency

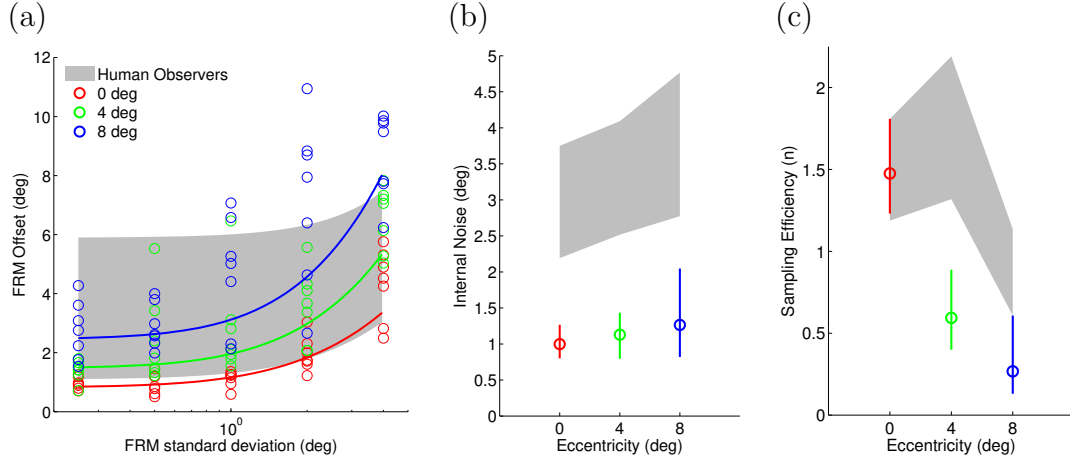


Figure 10.11: **Equivalent Noise Analysis on the model responses.** As Figure 10.10, except for the model. (a) Equivalent noise functions at fovea (red), 4 degrees (green) and 8 degrees (blue) eccentricity. Individual data points are discrimination thresholds for all seven separate model parameter configurations. Curves are best fitting equivalent noise functions to the data averaged across the seven model parameter configurations. Data is overlain onto a gray shaded region representing the region of observed human performance. (b) Internal Noise and (c) Sampling Efficiency parameters of the estimated equivalent noise functions as a function of eccentricity. Data are the average across the seven model parameter configurations. Error bars represent 95% bootstrapped confidence intervals. Gray shaded region represent the 95% bootstrapped confidence regions from the human observer data.

decreases at a faster rate in the periphery of the model than in the periphery of human observers.

I further evaluated the influence of the model's parameters on the model's performance on the EN task. Step-wise linear regressions were calculated to assess the relationship between the model's parameters (Table 10.1, predictors) and the Internal Noise and Sampling Efficiency (dependent variables) averaged across eccentricities for each parameter configuration. The step-wise regression found no significant relationship between any of the model's parameters and the model's estimated Internal Noise (no predictors were added to the constant regression model). A significant regression equation was found instead between the percent of the model's cortical area devoted to the fovea ( $\chi$  parameter) and the model's estimated Sampling Efficiency:  $F(1,5)=27.6$ ,  $p=0.003$ ,  $R^2 = 0.816$ . As can be seen in Figure 10.12, the greater the portion of the model's cortex was dedicated to the fovea, the worse the Sampling Efficiency. This may be due to the over-representation of the fovea in the model's cortex, i.e. using too many cortical processing units to represent the fovea leaves too few cortical units for the periphery. The decrease in Sampling Efficiency in the periphery of both model and humans is thus sensibly attributable to receptive field size changes across the visual field, which are in turn driven by the lossy log-polar mapping that projects the retinal image onto primary visual areas.

Experiments 1 and 2 found that human observers show a systematic foveocentric bias in FRM estimation, whereas the model does not exhibit such bias. Conventional wisdom would associate the bias with log-polar oversampling toward the fovea, thus the lack of foveocentric bias in the model was perplex-

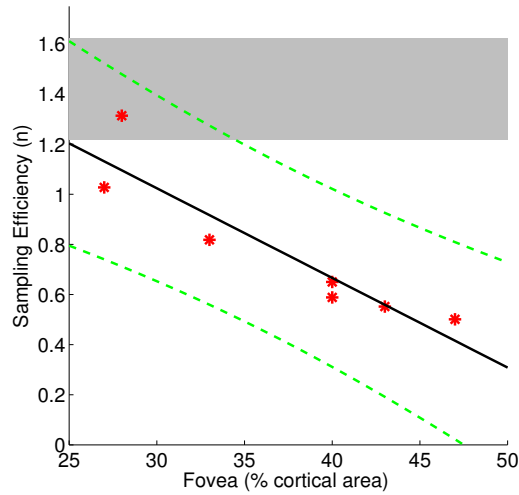


Figure 10.12: **Sampling Efficiency depends on the portion of the model’s cortical area dedicated to the model’s fovea (see Table 10.1).** Red data points are estimated Sampling Efficiency, averaged across eccentricities for each model parameter configuration, plotted against the percent of the cortical area representing the model’s fovea. Black line is best fitting linear regression line, bounded by 95% confidence intervals of the fit (green dotted lines). Gray shaded region represents 95% confidence bounds of the estimated Sampling Efficiency in human observers, averaged across eccentricities for each observer.

ing. However, having observed that the model has significantly less internal noise than human subjects, I asked whether foveocentric bias arises from log-polar sampling in noise. I employed the model's FRM estimates in Experiment 4 to compute the model's foveocentric bias as a function of external noise. Figure 10.13 confirms that at low levels of noise the model exhibits little or no foveocentric bias. However, I find that the model's foveocentric bias increases monotonically as a function of external noise. This suggests that the foveocentric bias observed in this study as well as in others (Johnston, White, and Cumming 1973; Warren Jr and Saunders 1995) is explained by log-polar oversampling toward the fovea in conditions of motion detector noise.

## 10.4 Discussion

In this chapter I have investigated a model of the human visual system that mimics the first processing stages of the dorsal cortical pathway for motion processing. The model was first developed by Solari, Chessa, and Sabatini (2014), where it was validated with regards to estimation of Time to Contact in real movies of approaching objects. The main feature of the model is that it functions on log-polar transformed image sequences, which can be thought of as cortical image sequences. Thus the model is able to characterize motion interpretation directly in the cortical domain. If the model were a good approximation of the human visual system, it would give us insight as to how the human visual system elaborates the distorted representation of the visual world that reaches cortical visual areas. The model is able to

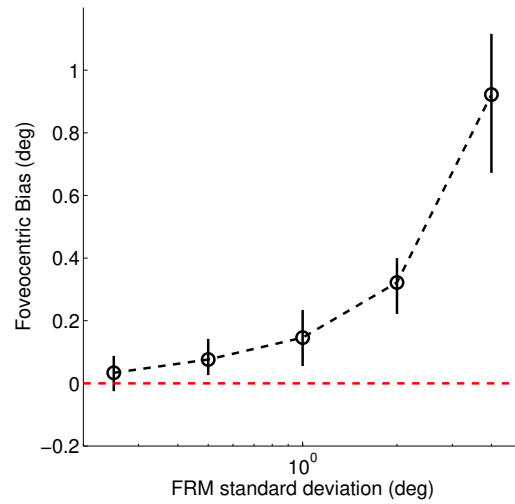


Figure 10.13: **Foveocentric bias of the neural model as a function of external noise.** Positive values of foveocentric bias mean that the model's FRM estimate was closer to the fovea than the true FRM location. Circles are the median bias at each level of external noise from all the data from Experiment 4. Error bars represent 95% bootstrapped confidence intervals. Dashed red line highlights the null level of foveocentric bias.

decompose optic flow information directly in the cortical domain, and could be a good candidate for neural computations leading to flow parsing (Foulkes, Rushton, and Warren 2013a), a necessary computational step in estimating object motion during self motion.

Few studies in the literature on cortical motion processing have addressed the log-polar transform. Grossberg, Mingolla, and Pack (1999) provide both a functional explanation and quantitative simulations of experimental data of MSTd cells. To accomplish this the authors analyze MST responses to inputs of optic flow patterns transformed into log-polar coordinates. Differently from the approach presented here, Grossberg, Mingolla, and Pack (1999) do not to mimic the entire visual pathway starting from images as inputs to the model. Thus the model investigated here extends previous work by describing how cortical optic flow maps can be obtained from the first stages of cortical processing, and then accounts for how this transformed cortical optic flow may be processed by areas MT and MSTd to compute FRM estimates. Furthermore, while Grossberg, Mingolla, and Pack (1999) indeed address the log-polar transform, the cortical mapping they considered is different from the one presented here. This did not allow Grossberg, Mingolla, and Pack (1999) to verify the effects of the different parameters of the mapping, in order to reach a cortical representation with performance similar to that of humans, as I have done in this chapter. Even though the model described here cannot be directly compared to that of Grossberg, Mingolla, and Pack (1999) since the two models require different inputs, it is interesting that both models provide similar insights, such as biases towards fixation for limited depth (see the findings regarding the foveocentric bias).



Another study in the literature that could be compared with the current approach is the ViSTARS neural model proposed by Browning, Grossberg, and Mingolla (2009). Both the ViSTARS model and the approach presented here similarly account for the performance of a non-foveated human observer. Browning, Grossberg, and Mingolla (2009) report errors of  $1 - 3$  deg in the estimation of the FRM with several kinds of visual stimuli. This result is comparable to the error of  $\approx 1$  deg, I observe near the fovea of both model and human observers. The main difference between the ViSTARS model and the model by Solari, Chessa, and Sabatini (2014) is that the ViSTARS model does not consider the log-polar mapping. Thus, the ViSTARS model cannot describe how optic flow computations and FRM estimates vary throughout the visual field. Conversely, the authors of the ViSTARS model investigate the effect of eye rotations on optic flow computations, whereas I do not address this issue.

One important aspect of the investigation presented in this chapter is that I aim to provide a direct comparison of the model's output with respect to the performance at FRM estimations of human observers. I employ the same set of naturalistic stimuli and tasks for both human observers and the model. In this way, I am able to test whether the model's behaviour is analogous to that of human subjects across the visual field. Whereas several works in the literature, e.g, Browning, Grossberg, and Mingolla (2009), compare their results with neurophysiological and behavioural data from the literature, none directly test their models on the same experimental procedures applied to human observers, nor do they test how varying the models' tuning parameters affects the models' performance.

In this chapter I summarize how to mathematically locate the FRM from the computed cortical optic flow. I then proceed to compare the model's performance on FRM estimation to the performance of human observers. Both model and human observers are able to reliably estimate the FRM in expanding and contracting naturalistic dead leaves stimuli (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001). The precision with which human observers and the model can estimate the FRM location worsens in the peripheral visual field. The magnitude and trend of observed results in the model is shown to be consistent with the pattern of results observed in human observers. The theoretically based neural computations implemented in the proposed model are thus a good candidate for the computations actually performed by the human visual system regarding complex motion estimation. The main factors contributing to the changes in the precision with which humans estimate the FRM throughout the visual field are resolution loss and receptive field size differences. Resolution loss and receptive field size changes throughout the visual field are also two of the main features of the implemented model. By employing EN analysis on both human observers and on the model, I show that increases in receptive field size across the visual field lead to decreases in Sampling Efficiency in the peripheral visual field of both human observers and the model. Receptive field size is in turn modulated by retino-cortical M-scaling. Thus, the retino-cortical transformation and the hierarchical architecture implemented in the model produce the same general effects as their biological counterparts in human observers. However, I did find some differences between the performance of the human observers and the model. The model's performance was better

than that of human observers for low levels of external noise, and worsened at a faster rate with increasing levels of external noise. Indeed the model had less Internal Noise as well as lower Sampling Efficiency with respect to human observers. The fact that the model had lower Internal Noise with respect to human observers is sensible: the model does not incorporate the various sources of Internal Noise that human observers are subjected to, such as the baseline level of random firing of cells, fixational eye movements and fluctuations in level of attention. Conversely, the fact that the model has lower Sampling Efficiency than human observers might reflect the fact that the model's internal parameters were heuristically selected from what was believed to be a sensible range. It might be interesting in future work to develop ways of training the model with natural stimuli and verify whether model parameters finely tuned to the spatio-temporal statistics of natural scenes provide patterns of performance more similar to those observed in human subjects. Future work should also focus on comparative simulations of the different models presented in the literature. Implementing the available models to incorporate space-variance across the visual field and tuning these models to perform as close to human observers as possible would allow researchers to determine the most parsimonious model that best accounts for human performance. This would provide further insight into the key processing stages underlying biological motion perception.

The results of the Equivalent Noise analysis differ from those reported by Bex and Falkenberg (2006). Human observer performance was overall higher in Bex and Falkenberg's study, their FRM thresholds were smaller by a factor of 4, Internal Noise was lower by a factor of 1.5, and Sampling

Efficiency was higher by a factor of 7. Bex and Falkenberg found that Internal Noise significantly increased with eccentricity, whereas I find a trend in that direction but it was not statistically significant. Furthermore, I find a significant decrease in Sampling Efficiency in the peripheral visual field, whereas they found the same trend, but it failed to reach significance. Unfortunately, individual estimates of Internal Noise and Sampling Efficiency are highly variable in both studies and are poorly constrained by the data. It is also possible that the differences between results may be driven by the different stimuli employed in the two studies. Whereas Bex and Falkenberg employed localized patches of random dot stimuli placed at various positions in the visual field, in the current study I employed full-field naturalistic dead leaves stimuli (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001). These stimuli have the same  $1/f$  spatial frequency spectrum and contrast range of natural images, and are textured with occlusions and edges at a variety of orientations. Hence the dead leaves are better suited to test the performance of the visual system under natural viewing conditions, since they better approximate the natural stimulus range in which the visual system operates.

The localized patches employed by Bex and Falkenberg might also have been ill suited to highlight the foveocentric bias at FRM estimation found in this and other studies (Johnston, White, and Cumming 1973; Warren Jr and Saunders 1995). The finding that the model's foveocentric bias increases with external noise suggests that this bias sensibly arises from log-polar sampling in noise.

With regard to computer vision applications, mimicking the log-polar

mapping adopted by foveated mammalian visual systems might provide interesting advantages. Log-polar mapping provides a wide field-of-view while maintaining high spatial resolution on the region of interest and thus providing significant data reduction. This could be a desirable feature in robotic active vision systems, where the fixation point of the cameras may continuously change, and a mechanism to obtain robust features by working on small images (since the cortical mapping also produces a consistent compression ratio, see Equation 10.3) would be useful to obtain real-time implementations (Berton, Sandini, and Metta 2006). In Solari, Chessa, and Sabatini (2014) and in Chessa, Maiello, Bex, et al. (2016), a general approach to extract visual features directly into the cortical domain is developed. Standard computer vision algorithms that work in the Cartesian domain can thus be employed directly in the cortical domain. In Solari, Chessa, and Sabatini (2014) the model was proven to reliably estimate time to contact in automotive real-world scenes. The presented model is thus a good candidate for robotics applications such as the humanoid robot iCub (Metta et al. 2008). Although locomotion might not solely be guided by optic flow information (Rushton et al. 1998; Warren Jr et al. 2001), and alternate design strategies are available for humanoid robotic locomotion (Rushton, Wen, and Allison 2002), optic flow is clearly at the basis of the perception of self-motion, object motion and visually guided action (Watson, Ahumada Jr, et al. 1985; Koenderink 1986; Orban et al. 1992; Perrone and Stone 1994; Simoncelli and Heeger 1998) and the model investigated in this chapter could be employed to further study and implement in robotic architectures the cognitive processes underlying biological motion perception and action.

## Part 4: Assessment of Low-Cost Technologies

# Chapter 11

## Evaluation of the Tobii EyeX Eye Tracking Controller and Matlab Toolkit for Research

### 11.1 Introduction

Eye-tracking technology provides a unique source of information about how humans and animals visually explore the world.

Until very recently, eye tracking technology has been prohibitively expensive for anything other than industrial, clinical, or well-funded basic research, with the cost of an eye tracker ranging up to tens of thousands of dollars. However, as consumer demand drives down the cost of new technology and increases its availability in our daily lives, so has eye tracking technology begun to be inexpensive. Minimal eye tracking systems are now being embedded in smart-phones, and low cost eye tracking devices are beginning to

appear on the market. Specifically two devices have attained the sub-\$150 price point: the EyeTribe (Dalmaijer 2014; Ooms et al. 2015) and the Tobii EyeX (Gibaldi et al. 2016).

This chapter is focused on reviewing the capabilities of the Tobii EyeX Controller, which is a low price, image based eye tracking device from Tobii AB, conceived for consumer applications. The claims put forth by Tobii AB are that this new device is designed for eye gaze interaction with natural user experience where the user can sit, stand and move around somewhat freely. The Tobii EyeX can be mounted on both desktop and laptop setups, allowing for immediacy and ease of use. Moreover, the eye tracker is advertised as not requiring regular re-calibrations and as being able to cope with a great variety of physiological factors such as eye color, ethnicity, refractive correction and age, independently of head movements and changing light conditions over time.

Tobii AB primarily produces research grade eye tracking devices. Although Tobii AB provides detailed technical specifications of all its research dedicated devices, an extensive description of the characteristics of the EyeX Controller is not available. Here I provide an empirical analysis of the characteristics and technical specifications of the device, in terms of accuracy and precision, latency and sample frequency. In order to enhance the usability of the device for research (*e.g.* Cornelissen, Peters, and Palmer 2002), I have contributed to the development of an open source Matlab Toolkit that can be used to interface with the eye tracker.



## The Tobii EyeX Controller

**Features of the System** The Tobii EyeX is an eye tracking device that allows moderately free head movements. It returns a real-time estimate of the left and right eye gaze positions on the screen, as well as the 3D position of the two eyes with respect to the screen center.

The actual technique exploited by the device for eye tracking is not declared by the manufacturer. Nevertheless, since the EyeX is based on Tobii's latest hardware, it is reasonable to assume that it relies on the same techniques employed by the other Tobii eye tracking devices (e.g. X2-60 or TX300). These eye trackers are based on the pupil center and corneal reflection technique. The position of the pupil (which moves jointly with the eye) is computed with respect to the position of a glint (which is relatively invariant of the movement of the eye) produced by an infra-red illuminator on the cornea. The angular difference between the position of the pupil and the position of the glint is used to estimate the eye gaze. To ensure robust estimates of pupil location, both bright pupil and dark pupil eye tracking techniques are employed. In bright pupil eye tracking, an illuminator is placed close to the optical axis of the imaging device, causing the pupil to appear lit up. During dark pupil eye tracking, the illuminator is placed away from the optical axis, causing the pupil to appear black.

The image processing necessary for gaze data calculations is performed by the Tobii EyeX Engine, that runs on the PC to which the device is connected via USB3. Multiple applications can be connected as clients to the Tobii EyeX Engine over a LAN connection. These applications can be employed to perform a calibration of the gaze data and to gather the eye gaze data

in real-time. The Tobii SDK released with the EyeX provides the libraries necessary to access the eye tracking data in the C/C++, C#/.NET, and Unity 3D programming languages. In order to enhance the accuracy of the gaze point estimation, the Tobii EyeX Engine provides a native calibration procedure (TNC) to be performed before the usage of the eye tracker by a new user. The procedure is required to compute the geometry of the setup (e.g screen size, distance, etc.) and to collect information about the light refraction and reflection properties of cornea of the subject.

Since the device is targeted for consumer applications, and not for scientific research, few technical specification have been provided by the manufacturer. However, when employing an eye tracker for scientific research, a precise characterization of the spatial and temporal performance of the device is essential: the accuracy and precision of the gaze estimation need to be evaluated, as well as the system latency and the sampling rate. **Gaze accuracy** refers to the average angular error in gaze estimation when a user is fixating a known location in space. **Gaze precision** refers to the spread of the estimates of angular gaze position when the eyes are steady and fixating a target. Since the eye tracker can potentially be employed for gaze-contingent applications, in which stimuli on a computer monitor change as a direct result of changes in gaze position, the **system latency** can be defined as the delay between a change in gaze location and the related change on the display. This end to end latency consists of the exposure time of the eye tracker camera, the image read-out and transfer time, the image processing time, the data transfer time between the Tobii EyeX Engine and the end application, and the display refresh rate. The **sampling rate** refers to the inverse of the

time interval between two consecutive eye position measurements. In order to validate the Tobii EyeX for scientific research, I propose and perform a series of procedures to provide a quantitative evaluation of the spatial and temporal characteristics of the device.

## **11.2 Methods**

### **Matlab Toolkit**

To allow for a broader and more direct use of the Tobii EyeX device in scientific research, I have contributed to the implementation of a software Toolkit in Matlab which interfaces with the eye-tracker controller. The Matlab Toolkit consists of four parts: 1) a client UDP (User Datagram Protocol) interface to connect Matlab with the Tobii server, 2) a set of basic connection functions for data transmission and reception, 3) a set of routines for standard use of the device, and 4) sample code provided to exemplify the usage of each function of the Toolkit in simple experiments. The graphical interface of the Toolkit has been implemented exploiting the Psychophysics Toolbox Version 3 (Brainard 1997; D. G. Pelli 1997). A detailed description of the Matlab Toolkit is provided in Gibaldi et al. (2016).

### **Experiment 1: Accuracy, Precision and Sampling Rate**

A very simple experiment was devised to assess the accuracy and precision and sampling rate of the EyeX device. Stimuli were presented on a 17 inch LCD with  $1920 \times 1080$  resolution at 60 Hz, run from an NVidia Quadro

600 graphics processing unit. Observers were positioned at 70 cm from the computer monitor with the head stabilized by a chin and forehead rest. The EyeX Controller was mounted at the bottom of the screen. Fifteen subjects participated in the experiment, all had normal or corrected to normal vision. Observers were shown a 0.5 degree dot target at each of the 25 locations shown in Figure 11.1 and were asked to maintain steady fixation on each target for 2 seconds. Each fixation location was shown four times to each subject.

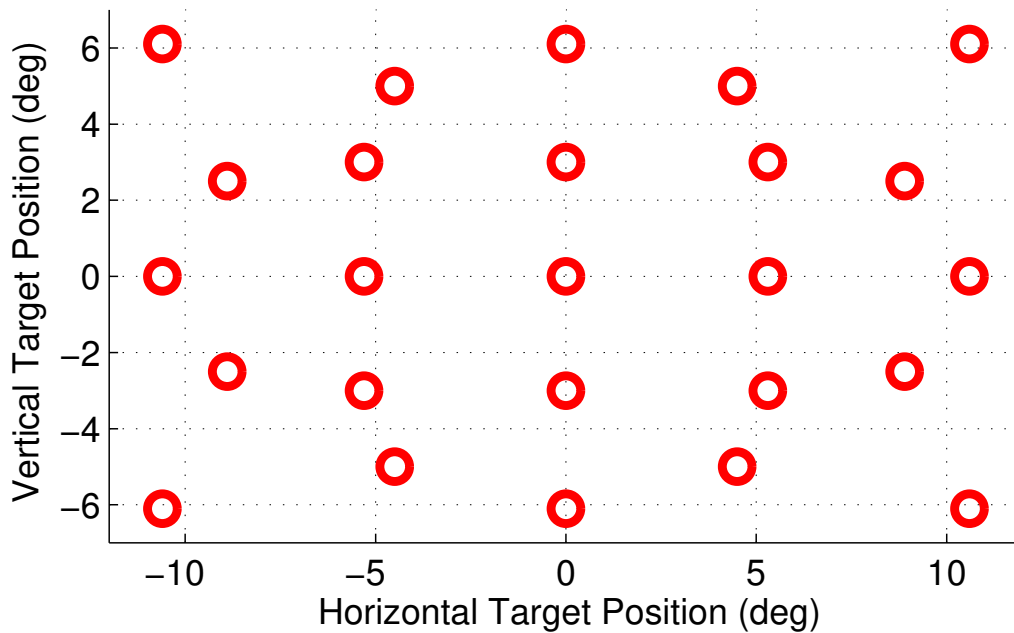


Figure 11.1: **Fixation locations selected to measure accuracy and precision.** (0,0) represents straight ahead fixation to the center of the monitor.

The EyeX outputs data samples which consist of the x and y on-screen coordinates for left and right eye paired to a time stamp specifying at what

time from the beginning of the eye tracking session the data sample was collected. The data collected during this simple experiment was employed to estimate the accuracy, precision and the sampling rate of the EyeX device. Gaze position during each fixation was computed by discarding the initial 0.5 seconds of data, which often correspond to compensatory saccades (Krauskopf, Cornsweet, and Riggs 1960; Cyr and Fender 1969), and then taking the median gaze position from the samples collected during each stimulus presentation. Accuracy was measured as the distance between the median gaze estimate and the true target location. Precision was computed as the standard deviation of the estimates of angular gaze position when the eyes were steady and fixating each target. Sampling rate was taken as the inverse of the time interval between adjacent samples.

## **Experiment 2: System Latency**

To evaluate the device latency for gaze contingent applications, a method similar to that described in Saunders and Woods (2014) was employed. A simple gaze contingent display, consisting of two fixed targets and a cursor under the control of the user's gaze, was implemented directly in C/C++ as well as with the Matlab Toolkit. The C/C++ gaze contingent implementation was compared against the Matlab gaze contingent implementation to assess whether the UDP server for data communication between Matlab and the Tobii EyeX Engine introduced any additional latency.

Observers were positioned at 70 cm from a 17 inch LCD with  $1920 \times 1080$  resolution at 60 Hz, run from an NVidia Quadro 600 graphics processing unit. Observers' heads were stabilized by a chin and forehead rest. The EyeX

Controller was mounted at the bottom of the screen. Observers were required to execute saccades back and forth between two targets presented on screen. Along with the saccade targets, the on-screen gaze position was displayed as a cursor in real time. While observers were performing the saccade task, a high speed camera was employed to record (at 240 fps) the PC screen and simultaneously the observer's eye through a mirror. Two observers performed 20 saccades each while the camera simultaneously recorded both their eyes as well as the gaze contingent cursor on the screen.

After having acquired these video sequences, a video editing program (VSDC Free Video Editor) was used to perform a frame by frame inspection of the video sequences. For each saccade executed by the subjects, the movie frame in which the eye movement was initiated and the movie frame in which the gaze-controlled cursor began to move across the screen were identified. The number of frames between the actual eye movement onset and the corresponding response of the on-screen cursor could then be unambiguously counted. The total latency with which the system responded to the eye-movement was measured by multiplying the number of elapsed frames by the duration of each camera frame (4.2 ms). The estimated latency thus resulted from the sum of the display latency (hardware) and the gaze computation latency (software).

### **Experiment 3: Saccades**

To bring our high resolution fovea onto targets of interest preselected with our low resolution peripheral vision, our oculomotor system continuously makes fast, ballistic eye movements called saccades. Saccades are perhaps the most

investigated type of eye movement, thus I devised a simple experiment to verify whether simple saccadic eye movements could be successfully measured with the EyeX.

Stimuli were presented on a 28 inch Dell P2815Q monitor with  $1920 \times 1080$  resolution running at 60 Hz, run from an AMD Radeon HD 7000 graphics processing unit. Observers were positioned 50 cm from the monitor, which subtended  $70 \times 40$  degrees of visual angle. Observers were positioned in a chin and forehead rest to stabilize head movements. The EyeX eye tracker was positioned below the monitor in front of the observers. Observers were instructed to fixate a central red fixation dot presented on a uniformly black screen, and when ready, were required to initiate a trial by pressing a key on the keyboard in front of them. The fixation target would then turn white, and, after a 500 ms delay, the target would jump 10 degrees left. Observers were simply required to visually track the target as accurately as possible. The target would remain at the eccentric position for 750 ms, and then turn red once again and return to the center of the monitor. Each subject performed 50 eye movement trials. Three subjects participated in the experiment, all had normal or corrected to normal vision.

## **Experiment 4: Smooth Pursuit**

Another commonly investigated class of eye movements are smooth pursuit eye movements, which allow us to closely track moving objects. I thus set out to verify whether we could reliably measure smooth pursuit eye movements with the Tobii EyeX in another simple experiment.

As in the previous experiment, stimuli were presented on a 28 inch Dell

P2815Q monitor with  $1920 \times 1080$  resolution running at 60 Hz, run from an AMD Radeon HD 7000 graphics processing unit. Observers were positioned 50 cm from the monitor, which subtended  $70 \times 40$  degrees of visual angle. Observers were positioned in a chin and forehead rest to stabilize head movements. Observers were instructed to fixate a central red fixation dot presented on a uniformly black screen, and when ready, were required to initiate a trial by pressing a key on the keyboard in front of them. The fixation target would then turn white, and, after a 500 ms delay, the target would begin to move at a constant speed of 10 degrees/second to the right. After one second, the direction of the target would reverse and the target would return to the center of the monitor. Observers were simply required to visually track the target as accurately as possible. Once the target had returned to the starting position, it would turn red and a new trial could be commenced. Each subject performed 50 eye movement trials. Three subjects participated in the experiment, all had normal or corrected to normal vision.

## **Experiment 5: Vergence**

When looking at an object binocularly, our two eyes must rotate in opposite directions to be correctly pointed towards the object. These disconjugate rotatory movements are called vergence eye movements. Vergence eye movements correctly position the retinal areas with highest spatial resolution of both eyes (the foveae) onto the object of interest, and thus facilitate binocular fusion, resulting in a richer perceptual experience of the selected object. Vergence eye movements are another commonly investigated class of eye movements. Thus I designed an experiment to evaluate the usability of



the Tobii EyeX in oculomotor research involving eye vergence.

Observers were positioned in a chin and forehead rest to stabilize head movements, at a distance of 100 cm from the screen, *i.e.* at a vergence distance of  $\sim 3$  degrees. Whereas the eye movement measurements described above could be performed using a conventional 2D monitor, the test of vergence eye movements required three-dimensional stimulus presentation. Accordingly, the experiment was conducted with a passive stereo LCD (LG 42LW450A) running at 100 Hz from an NVidia Quadro 600 GPU. Observers were required to wear stereoscopic polarized glasses.

The size of the employed screen (42") was larger than the screen size (24") suggested by the manufacturer of the EyeX. However, eye tracking was still possible simply by placing the eye tracker on a stand at 60 cm from the observers. To obtain reliable gaze data the device had to be positioned parallel to the screen, as if it were mounted at the bottom of the display.

The visual stimulus employed to drive binocular fusion was a flat virtual plane positioned in the center of the screen. The stimulus subtended 10 degrees of field of view to ensure full coverage of the area of the field of view that elicits vergence movements (Allison, Howard, and Fang 2004). The plane was textured with 1/f pink noise, which has the same frequency content of natural images (Kretzmer 1952; Bex and Makous 2002; Jansen, Onat, and König 2009). A white fixation cross was presented in the center of the stimulus.

The stimulus protocol was conceived to test both divergence and convergence eye movements. The plane was initially presented with 1 degree of positive disparity, thus requiring observers to fixate at a vergence distance

of 4 degrees. Once a subject was properly fixating (which took  $\sim 2$ s), the stimulus disparity was set to zero, *i.e.* the plane would be rendered at the actual depth of the screen, thus inducing a divergence movement. This procedure was repeated 50 times, and alternated with a -1 degree disparity step, which required a convergence movement. Three subjects participated in the experiment, all had normal or corrected to normal vision.

## 11.3 Results

### Experiment 1: Accuracy, Precision and Sampling Frequency

From the data collected in Experiment 1, the observed sampling time of the system was  $18.05 \pm 2.49$  ms (median  $\pm$  inter quartile range), resulting in a median sample frequency of  $\sim 55$  Hz.

The performance of eye tracking devices may vary as a function of gaze angle away from straight ahead, central fixation. The data collected in Experiment 1 were thus separated with respect to the eccentricity of the visual target, computed as its angular distance from the center of the screen. This resulted in eight values of eccentricity, ranging from 0 degrees to  $\sim 12.2$  degrees. The angular error did not follow a Gaussian distribution, but was better described by a Poisson error distribution. Thus, rather than employing mean and standard deviation, performance metrics are reported in terms of median and inter-quartile range. Accuracy is reported as the distance between the median gaze estimate and the true target location. Precision is

computed as the standard deviation of the estimates of angular gaze position when the eyes are steady and fixating a target.

Figure 11.2 summarizes the results obtained regarding accuracy (a) and precision (b) of the Tobii EyeX as a function of visual angle. The device performs best at the center of the display. Accuracy worsens slightly at increasing eccentricities, whereas precision is approximately constant (*cf.* the linear regression lines). Accordingly, near the center of the monitor accuracy and precision can be considered to be  $< 0.4$  degrees, and  $< 0.2$  degrees respectively. At more than 5 degrees away from the center of the monitor, accuracy and precision worsen to  $< 0.6$  degrees, and  $< 0.25$  degrees respectively.

## Experiment 2: System Latency

The latency estimated from the data collected on two observers with the C++ gaze-contingent implementation was  $48 \pm 3$  ms (mean  $\pm$  standard deviation). The latency observed with the Matlab Toolkit gaze-contingent implementation was  $47 \pm 4$  ms. These data confirm the reliability of the procedure employed to estimate latency, since the uncertainty on the latency estimates is primarily due to the temporal resolution of the camera (4.2 ms). Although different total latencies may be possible with different display or PC configurations (Saunders and Woods 2014), these data show that the UDP communication link between the Tobii server and Matlab does not appear to influence the system latency.

Because saccadic suppression (Volkman 1962; Volkman, Schick, and Riggs 1968; Volkman et al. 1978) or poor sensitivity to high speed retinal

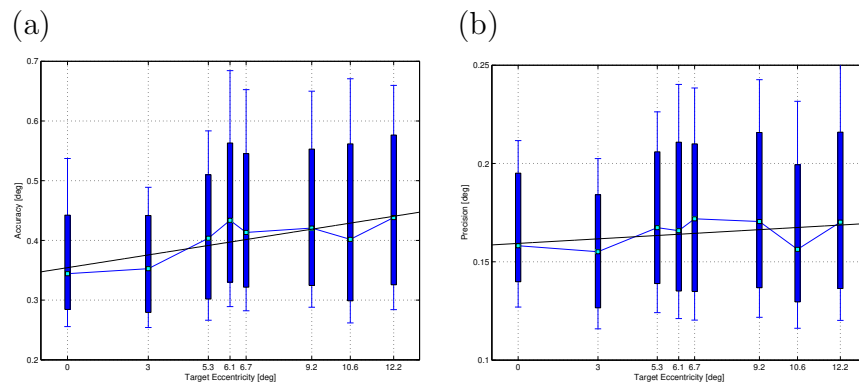


Figure 11.2: **Accuracy and precision of the Tobii EyeX.** Distribution of Accuracy (A) and Precision (B) of the Tobii EyeX as a function of target eccentricity. Data are aggregated from across the fifteen observers. Green squares represent the median values, thick blue bars represent inter-quartile range, blue whiskers encompass minimum and maximum observed values, and the black lines are the linear regression lines passing through the median values.

images (Dorr and Bex 2013) render a person visually insensitive for about 50 ms from the beginning and end of a saccade, the observed system latency is on the edge of noticeability for human users employing the system for gaze contingent applications.

These measurements regarding latency and sampling frequency are necessarily system dependent. Thus, as a final consideration, it is worth noting that the use of a high performance PC and low-latency monitor are likely to improve the overall performance of the eye tracking system.

### **Experiment 3: Saccade Dynamics**

Figure 11.3 shows the results from Experiment 3 regarding the measurements of saccade dynamics in three observers. The first subject was an experienced observer (author GM), while second and third subject were naive observers. Figures 11.3a-c present average horizontal eye position as a function of time from target step for the saccades measured in all three subjects. As can be seen from the shaded regions representing the variability in the measurements, the data collected on the first two subjects (Figure 11.3a,b) were highly reliable and accurate, whereas the data collected on the third subject (Figure 11.3c) were more variable and particularly less accurate for the subject's right eye (red trace) than for the subject's left eye (blue trace). The saccades in all three subjects were initiated between 200-250 ms after the onset of the eccentric target, which is consistent with typical saccade latencies observed in the literature (Saslow 1967; Cohen and Ross 1977). The duration of the saccades was  $\sim 50$ ms, which is also highly consistent with the literature on similarly sized saccades (Baloh et al. 1975; Bahill, Brockenbrough,

and Troost 1981; Behrens, MacKeben, and Schröder-Preikschat 2010).

Saccade velocity and saccade acceleration profiles are eye movement characteristics often investigated in the literature. I measured velocity (Figure 11.3d-f) and acceleration (Figure 11.3g-i) by taking the first and second derivative of the data in Figures 11.3a-c using a two point differentiator. Qualitatively, reasonable velocity and acceleration profiles are observable in all subjects. Peak velocity was  $\sim 400$  degrees/second, whereas peak acceleration and deceleration were  $\sim 18000$  degrees/second<sup>2</sup>, all values highly consistent with previous measurements of these parameters in normally sighted subjects (Bahill, Brockenbrough, and Troost 1981).

**Experiment 4: Smooth Pursuit Eye Movements** Figure 11.4 shows the results of the measurements of smooth pursuit eye movements in Experiment 4. As in the saccade experiment, the data collected on the first two subjects (Figure 11.4a,b) were highly reliable and accurate, whereas the data collected on the third subject (Figure 11.4c) were more variable. The typical characteristics (Robinson 1965; Spering and Montagnini 2011) of smooth pursuit eye movements can nonetheless be clearly observed in the data from all three subjects. In the initial *open-loop* stage of the tracking eye movement, after a latency ranging from 100-300 ms, the eyes accelerate and perform catch up saccades to capture the target. Then, in the *closed-loop* phase of the tracking eye movement, the eyes of the observers match the position of the moving target quite closely by maintaining the same speed as the target. When the target abruptly changes direction of motion, once again the eyes of the observers catch up and then match the smoothly moving target.

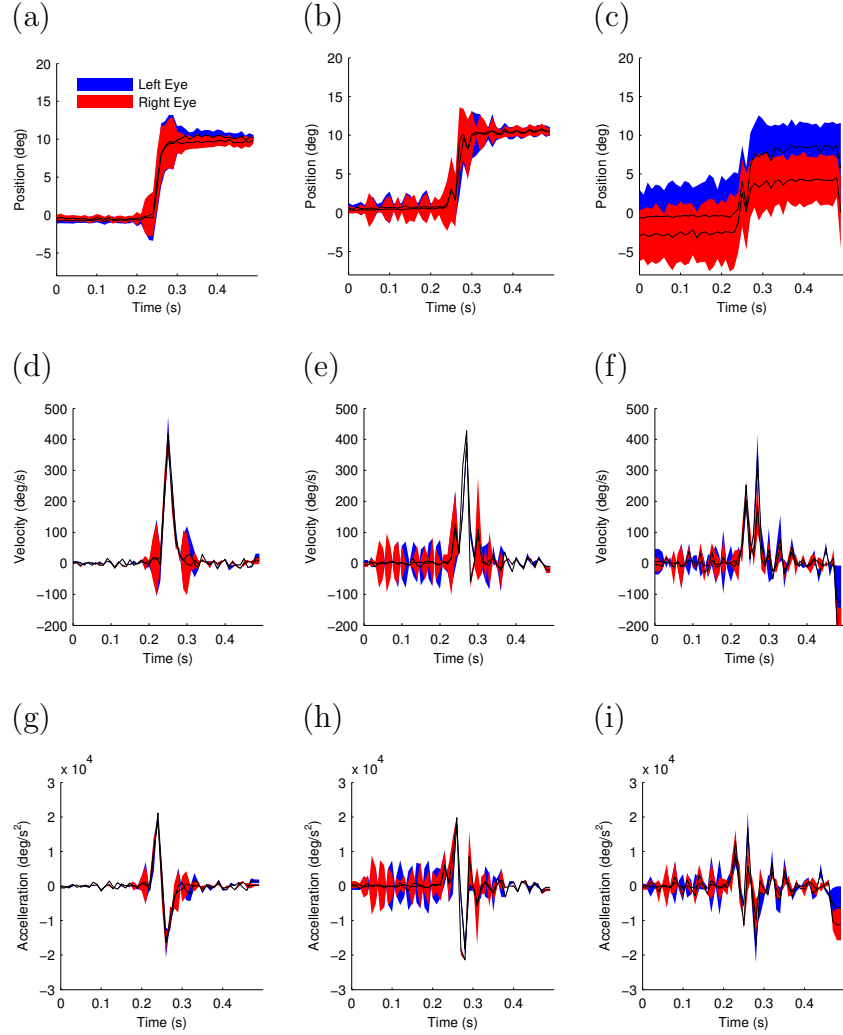


Figure 11.3: **Stereotypical saccade dynamics.** Saccade dynamics measured for left (blue) and right (red) eye in three subjects (columns) for 10 degree horizontal saccades. (a-c) Horizontal eye position as a function of time from saccade target onset. (d-f) Horizontal eye velocity. (h-i) Horizontal eye acceleration. Data are the average from 50 trials. Shaded region represent  $\pm 1$  SD.

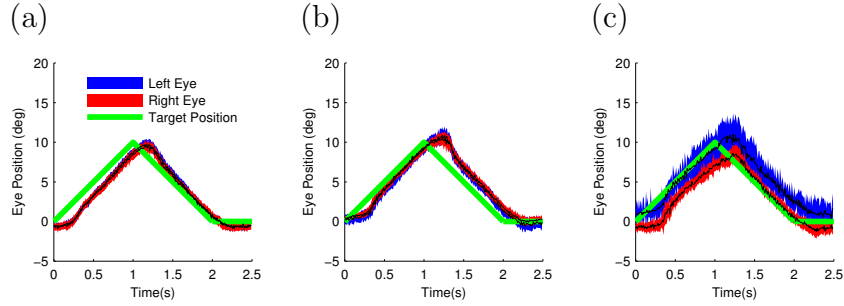


Figure 11.4: **Stereotypical pursuit traces.** Pursuit eye movements measured for left (blue) and right (red) eye in three subjects (a-c) for smooth eye movements in pursuit of a target (green trace) moving at 10 degrees/second from the center of the screen to ten degrees right of center and back. Data are the average from 50 trials. Shaded region represent  $\pm 1$  SD.

## Experiment 5: Vergence

Figure 11.5 shows the results of the measurements of vergence eye movements in Experiment 5. The first subject was an experienced observer, while the second and third subjects were inexperienced naive observers. Qualitatively we can observe from Figure 11.5a-c how the device provides a reliable characterization of the vergence trajectories. The eye movement response delay from stimulus onset was between 100 – 200 ms, whereas the time required to complete the movement was around 400 – 500 ms, which is all in good agreement with the literature (*e.g.* Hung et al. 1994; Collewyn, Erkelens, and Steinman 1995; Alvarez et al. 2002). As per the data collected on saccadic eye movements, I measured velocity (Figure 11.5d-f) and acceleration (Figure 11.5g-i) by taking the first and second derivative of the data in Figures 11.5a-c using a two point differentiator. Peak velocity was recorded at



3 – 5 degrees/second, while time to peak velocity was between 400 – 550ms. The measurements regarding acceleration were noisy, but qualitatively the expected patterns were observed.

## 11.4 Discussion

In this chapter I have presented qualitative and quantitative analyses of the characteristics and technical specifications of the Tobii EyeX Controller for its possible use in research applications. I have quantified accuracy, precision, latency and sampling frequency of the device.

Table 11.1 presents a comparison between the performances of the EyeX and other eye tracking devices at various price points. The technical specifications reported for the EyeX are those measured in this study. The technical specifications reported for the other eye trackers are taken from the specification sheets provided by the manufacturers.

The accuracy of the EyeX is comparable to that of both low and high-end devices. Conversely, the observed precision of the EyeX device is worse than any of the values reported by the manufacturers of the other devices. The system latency of a gaze contingent display implemented with the Tobii EyeX ( $< 50\text{ms}$ ) is comparable to the the system latency measured with research grade eye trackers (Saunders and Woods 2014) and is acceptable for at least some gaze contingent applications, such as gaze-contingent multiresolutional displays (Loschky and Wolverton 2007). The main difference between the Tobii EyeX Controller and research grade eye tracking technology is the sampling frequency. The sampling rate of Tobii EyeX was measured at  $\sim 55$

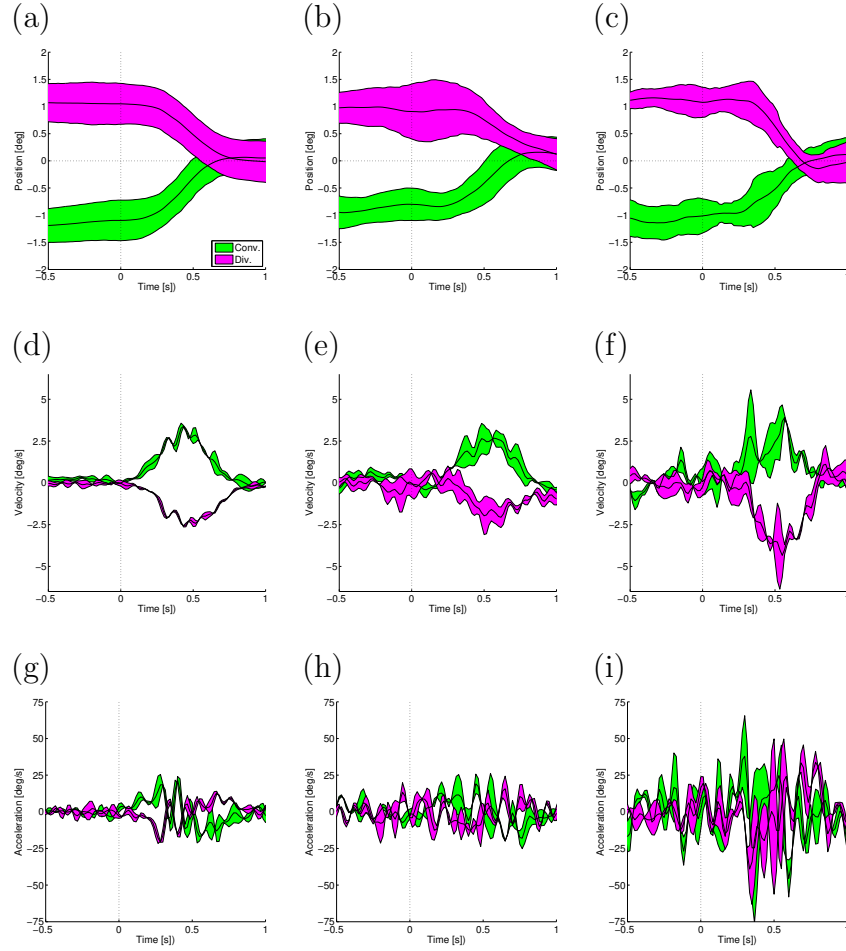


Figure 11.5: **Stereotypical vergence dynamics.** Vergence dynamics measured for convergence (green) and divergence (pink) eye movements in three subjects (columns) for  $\pm 1$  degree of vergence demand. (a-c) Vergence position as a function of time from vergence target onset. Zero vergence represents the actual depth of the screen. (d-f) Vergence velocity. (h-i) Vergence acceleration. Data are the average from 50 trials. Shaded region represent  $\pm 1$  SD.

Eye Tracker	Accuracy [deg]	Precision [deg]	Sampling Rate [Hz]	Latency [ms]	Price Point [\$]
<b>EyeX</b>	0.5-1	0.25	55	<50	~ 100
<b>EyeTribe</b>	0.5-1	0.1	30-60	<20	~ 100
<b>GP3</b>	0.5-1	0.1	60	<50	<1000
<b>myGaze</b>	0.5	0.1	30	<50	<5000
<b>SMI-REDm</b>	0.5	0.1	60-120	<20	<25000
<b>ViewPoint</b>	0.25-1	0.15	90-220-400	<10	<25000
<b>EyeLink 1000</b>	0.25-0.5	0.01	250-500-1000-2000	<10	>25000
<b>Tobii TX300</b>	0.3-0.8	0.1	60-120-250-300	<10	>25000

Table 11.1: **Commercial eye tracker comparison.**

Hz with our setup. Research grade eye trackers instead provide sampling frequencies up to 2000 Hz (*e.g.* the EyeLink 1000 with 2000 Hz camera upgrade).

Simple eye movement experiments have shown that the Tobii EyeX can be successfully employed to measure saccadic, smooth pursuit, and vergence eye movements.

This evaluation demonstrates that the EyeX is a potentially useful device for multiple research applications. Specifically, this device may be well suited for applications such as fixation compliance and monitoring of simple eye movement parameters. The EyeX has in fact so far been successfully employed for fixation compliance in an array of experiments regarding the measurement of: contrast sensitivity, letter acuity and crowding (Maiello, Harrison, et al. 2015; Carroll et al. 2016); motion discrimination (Maiello et

al. 2015a; Chessa, Maiello, Bex, et al. 2016); reading speed (Bex, Ayeni, and Wiecek 2015); illusory object completion (Ayeni, Harrison, and Bex 2015); retinal disparity patterns experienced by an observer (Gibaldi, Canessa, and Sabatini 2015); active binocular fixation strategy in 3D environments (Gibaldi, Canessa, and Sabatini 2016). These data have been successfully collected from expert psychophysical observers, undergraduate students, and even clinical populations. An intriguing possibility would be that of employing low cost eye trackers such as the EyeX in continuous target-tracking tasks to rapidly measure visual function (Bonnen et al. 2015). The small dimensions and portability of the device also make it a good candidate for field experiments where large and expensive devices (such as the EyeLink which requires a dedicated PC) are not easily employed.

However, while the low-cost nature of this device makes it an optimal candidate for gathering preliminary data and pilot testing novel ideas, the low sample frequency and limited precision of the device are not yet sufficient for all research applications. The temporal resolution of the EyeX is clearly insufficient to study perisaccadic visual perception (Ross et al. 2001). The precision of the device is also unlikely to be sufficient in measuring fine oculomotor adjustments such as those observed in saccade adaptation paradigms (Pélisson et al. 2010). Clearly the measurement of tiny microsaccadic eye movements (for a recent review see Rolfs 2009) is well beyond the capabilities of the device. Nevertheless, the Tobii EyeX, together with other emerging low-cost devices (Dalmaijer 2014; Ooms et al. 2015) and recent developments in web-cam based eye tracking (Xu et al. 2015), represents a meaningful step towards a widespread adoption of eye tracking technology,

both in commercial and research applications.

## Chapter 12

# The Perceptual Quality of the Oculus Rift for Immersive Virtual Reality

### 12.1 Introduction

Technical developments in the world of gaming and entertainment have been fostering the exploration of novel human-computer interactive methods and techniques. Stereoscopic 3D technology, the Microsoft Kinect and the Oculus Rift devices are examples of technologies initially developed for gaming and soon applied to other contexts such as human-computer interaction (HCI) research, museum installations, education and training, edutainment (i.e. the union between entertainment and education), virtual rehabilitation and surgical procedures (Bernhaupt, Isbister, and De Freitas 2015).

A well established field of application of Virtual Reality (VR) is to help

individuals who are anxious in certain social situations overcome their anxieties (North, North, and Coble 1996; Kwon, Powell, and Chalmers 2013). This kind of intervention is usually called virtual reality exposure therapy (VRET), and it is based on the generation of immersive virtual worlds, which are able to safely simulate situations that provoke anxiety in users by integrating real-time graphics, visual displays (*e.g.* 3D displays or head-mounted displays), body trackers, and other sensory input devices (Krijn et al. 2004). The use of VRET has several advantages, including the fact that subjects can be placed in simulated anxiety-provoking situations in a controlled manner, and health care professionals can gradually tune the different (virtual) situations to the patient’s state of progress (Botella et al. 2004). Other promising clinical uses of virtual reality include rehabilitation after traumatic brain injury (Grealy, Johnson, and Rushton 1999) or stroke (Laver et al. 2012), and novel treatments in amblyopia therapy (Vedamurthy et al. 2016).

In order to create environments that may be successfully used in contexts like VRET or edutainment, VR technology must be sufficiently immersive and must correctly stimulate the senses and the emotions of the users (Krijn et al. 2004). Moreover, VR environments should induce in users the sensation of presence (*i.e.* the sensation of physically being in the virtual environment). In the literature, the topic of presence is often addressed together with the topic of immersion (Cummings and Bailenson 2016).

Nowadays, creating and using virtual environments is possible with a variety of different technologies. The VR technologies most frequently employed in the literature are (large) 3D monitors, head-mounted displays (HMDs) and immersive rooms (Bouchard 2011). Large monitors (specifically stereoscopic

3D monitors) may provide a compelling yet limited immersive experience and until recently were the least expensive solution. HMDs and immersive rooms both provide a more powerful feeling of being immersed in the virtual environment. HMDs typically consist of goggles with small monitors mounted in front of each eye and a motion tracking system that monitors the position and movements of the user's head. In the literature, many authors analyze the issues of misperception that affect HMDs (Sharples et al. 2008; Ukai and Kibe 2003; Ukai and Howarth 2008). Amongst the wide range of commercially available HMDs only a few are characterized by both affordable costs and good performances. The Oculus Rift, which is the HMD device studied in this chapter, is available for \$599 and claims very high performance. A very different solution to create immersive VR environments is the immersive room, in particular the C-Automatic Virtual Environment, better known with its trademark name CAVE (Cruz-Neira, Sandin, and DeFanti 1993; W.-J. Li et al. 2001). The main drawback of this technology is that it consists of an expensive and room-size system, thus it cannot be easily employed in everyday or clinical use.

Another fundamental problem that must be considered when creating VR immersive environments, and that represents one of the main issues of such kind of systems, are adverse symptoms that may arise from VR use. Several works in the literature analyze the main drawbacks affecting stereoscopic 3D technology. Kooi and Toet (2004) experimentally determined the level of discomfort experienced by an observer with a wide range of imperfections and distortions commonly present in stereoscopic rendering technology (for in depth reviews regarding discomfort with stereoscopic image content



see Ukai and Howarth 2008; Lambooij et al. 2009; Bando, Iijima, and Yano 2012). A quantitative model describing the misperception problems caused by viewing a stereoscopic content from the wrong distance is presented in Held and Banks (2008). Recently, Hands, Smulders, and Read (2015) presented a study about distortions that arise when viewing stereoscopic screen from oblique angles. A serious but still unsolved issue that affects stereoscopic 3D monitors is the vergence-accommodation conflict. It is difficult to render a large range of 3D space without causing oculomotor stress to the observer, since the his/her eyes have to maintain accommodation onto the display screen (i.e., at a fixed distance) while also verging away from the display screen. Stereoscopic rendering technology thus breaks the natural relationship between accommodation and vergence eye movements (Shibata et al. 2011; Rushton and Riddell 1999; Wann, Rushton, and Mon-Williams 1995), and few attempts have been made to build displays that minimize this conflict (D. M. Hoffman et al. 2008; Love et al. 2009; Liu and Hua 2010; Johnson et al. 2016).

Other potentially adverse symptoms may be even more severe than those induced by rendering distortions and the vergence-accommodation conflict. In the real world, when a person moves, e.g. he/she changes the position of their eyes or head, the projections of the 3D real world immediately shift on the retinas, and at the same time the vestibular system indicates movement of the head. Due to hardware and software limitations, in CAVE and HMD VR systems there is an unavoidable delay between a user's movements and the updating of the virtual rendered scene. If this delay is excessive, the sensory information from the user's visual and vestibular systems might be

conflicting, and this can result in symptoms such as nausea, stomach awareness, dizziness, and headache. The sickness experienced in such cases may be referred to as “simulator sickness” or “cybersickness”. Simulator sickness and cybersickness are actually two slightly different phenomena. The causes, methods of measurement, and nuanced differences between simulator sickness and cybersickness are addressed by a recent review (Davis, Nesbitt, and Nalivaiko 2014) in which the authors analyze the use of questionnaires as a way of determining participants experience and susceptibility to VR and conclude that more cost-effective and objective physiological measures still need to be developed.

Given the widespread diffusion of new and inexpensive devices originally designed for games and entertainment, it is of general interest to test whether these devices can be effectively used in contexts different from the ones for which they were designed. As discussed above, some typical applications of VR devices include cognitive or physical rehabilitation and treatment of patients with phobias. These applications require that the devices should be immersive enough, i.e. the user or patient should perceive the virtual environments as realistic and they should interact with the VR environments in a natural way. Moreover, the devices should not provoke in the user undesired effects such as cybersickness or visual fatigue.

The aim of this chapter is to investigate whether the Oculus Rift head-mounted display can generate a perceptual experience similar to that experienced in the real world. One potential long-term goal of these investigations could be to use of the Oculus Rift and other HMDs for psychological rehabilitation, e.g. in the field of phobia treatments. It is worth noting that the

use of the Oculus Rift in clinical applications is recent, the first clinical data relating to medical applications of the Oculus Rift have appeared in 2014 (H. G. Hoffman et al. 2014). A recent study concerning the visual-vestibular interactions in self-motion perception by using the Oculus Rift is presented in Kim et al. (2015). However, clinical applications are out of the scope of this work. Here, I devise and validate a series of tests and measures to verify the perceptual quality of virtual reality scenarios experienced through the Oculus Rift. I further compare the perceptual quality of the Oculus Rift against the perceptual quality of the Google Cardboard and a widescreen 3DTV.

More specifically, this chapter addresses the following two issues about the perceptual quality of the Oculus Rift.

- i Immersivity. Is the Oculus Rift able to make the user feel like he/she was in a real world scenario? Is it possible to elicit in the user the sensation of presence via the virtual stimuli rendered by the device?

Both subjective and objective measurements were employed to assess the level of immersivity (or feeling of presence, see Waterworth et al. 2010) provided by the Oculus Rift. Specifically, observers' heart rate and head movements were monitored during VR experiences, and observers were asked to compile specifically developed questionnaires regarding the feeling of immersion during the VR experiences. The monitoring of physiological activation, including heart rate monitoring, is a commonly employed technique to measure levels of anxiety (Bradley and Dunlop 2005; Bailenson et al. 2007; Gorini et al. 2011). The measurement of observers' head movements in correspondence of virtual

obstacles was designed to verify whether observers experienced virtual obstacles as sufficiently looming to require moving out of the way (Kim et al. 2015). Multiple questionnaires have been developed in the literature to assess user experience in VR, but none are widely adopted (Gorini et al. 2011; Kwon, Powell, and Chalmers 2013). Thus, specific questionnaires were developed, tailored to the current investigation.

- ii Cybersickness. Does Virtual Reality experienced through the Oculus Rift induce physical discomfort to the user?

To answer this question, observers were administered the standard Simulator Sickness Questionnaire (Kennedy et al. 1993).

## 12.2 Methods

### Apparatus

Experiments 1 and 2 were carried out with the Oculus Rift DK2. The DK2 is the developer kit version 2, which provides a more advanced version of the device compared to the previous DK1. The device has a display resolution of 960 x 1080 pixels per eye, with a refresh rate of 75 Hz, and nominal horizontal and vertical field of view of 100 degrees. The displays use the low persistence OLED technology. Persistence is the amount of time pixels remain illuminated during each frame. In the experiments described below persistence was set to 2 ms to reduce motion blur. The Oculus Rift device is also equipped with a gyroscope, an accelerometer, and a magnetometer to track the head position. A near Infrared CMOS Sensor provides additional

position tracking.

In Experiment 3 the Oculus Rift was compared to two low-cost VR systems: a Google Cardboard and a 3DTV. Google Cardboard is a VR platform developed by Google for use with a head-mounted mobile phone. Its name comes from the foldout cardboard viewer which can be bought online for \$20. The Google Cardboard platform does not however require a specific smartphone mount; users may either build their own viewer from simple, low-cost components using specifications published by Google, or purchase a viewer manufactured by a third-party. In this work, I have used a compatible cardboard mount manufactured by Afunta which conveniently included an adjustable head strap. The mobile phone employed for the Google Cardboard system was a Samsung Galaxy S5 smartphone running on Android 5.0 with a 5.1-inch (13 cm) AMOLED display running at 60 Hz with 1920 x 1080 resolution. The lenses of Google Cardboard devices are designed for an 80 degree circular total field of view.

The 3DTV system employed was a 47-inch (119.4 cm) LG 47LA6200 passive 3D TV running at 120 HZ with 1920 x 1080 resolution. The TV renders stereoscopic 3D using FPR (Film-type Patterned Retarder) technology, a passive 3D system based on circular polarization. Subjects were placed approximately 1 meter from the screen, thus the monitor subtended approximately 50 x 35 (HxV) degrees of visual angle.

In the experiments described below the Oculus Rift HMD, the Google Cardboard and the 3DTV were connected to a PC running on the Windows 7 OS with an Intel Core i7 processor, 8GB RAM, and an Nvidia Quadro K4200 graphic card. The different virtual scenarios employed in this investi-

gation were created using Unity 3D cross-platform game engine. The same VR content created for the Oculus was rendered on the Google Cardboard system using TrinusVR, a software application capable of outputting VR content running on a PC onto any mobile device. The Intermed SAT-200 pulse oximeter was employed to record observers' heart rate in Experiment 1. In Experiment 3 instead heart rate was recorded using a HS-2BT heart rate sensor connected via Bluetooth to a Samsung Galaxy S5 smartphone running on Android 5.0. The accelerometer integrated within the Oculus Rift was employed to record users' head movements in Experiment 2. A Microsoft Kinect (version 1) was employed to measure users' head movements in Experiment 3 by recording the skeleton joint positions provided by the Microsoft Kinect SDK. Data analysis was performed using Matlab version R2014b (Mathworks).

## **Stimuli**

Four different virtual scenarios were employed in the current investigation (12.1). Three scenarios were developed ad hoc for this study, and one scenario was selected from the demos available from the official Oculus online forums. All scenarios contained stereoscopic depth information with appropriate disparity variation.

- i train scenario (developed ad hoc): the observer is placed onto a small railroad wagon. The wagon runs on rails along a circular path. Part of the path is placed within a wooded area and part of the path goes through rocky tunnels (see Figure 12.1a). The scenario is approximately 50 seconds long. The main purpose of this scenario was to

place subjects within a potentially immersive yet simple virtual environment. The scenario was employed to verify whether even a simple virtual environment would be judged immersive and be sufficient to induce a physiological reaction (measurable with the pulse oximeter) in healthy subjects.

ii skyscraper scenario (developed ad hoc): the observer is initially placed onto an open air elevator at the ground floor of a skyscraper. The skyscraper is placed within a city. The building, on which the user stays, is surrounded by smaller buildings and streets as well as multiple skyscrapers in the distance. The elevator rises to the top of the building. When the elevator stops, the subject is free to look around for a few seconds. The subject is then placed at the ledge of the building and instructed to look downwards. Finally, the observer is abruptly dropped to the ground floor (see Figure 12.1b). The scenario finishes with the observer crashing into the ground. The duration of the scenario is variable due to the interaction between the observer and the scenario itself, but the scenario lasts for approximately 1-2 minutes. This scenario was developed to assess self-reported experience and physiological reaction of healthy subjects immersed in a virtual scenario that could induce vertigo.

iii rollercoaster scenario: this scenario, known as the RiftCoaster HD, was professionally developed using the Unreal Engine 4 and is widely popular within the Oculus developer community. The scenario is available for download at <https://share.oculus.com/app/riftcoaster-hd>. This sce-

nario consists of a rollercoaster tour through a castle placed in a mountainous region, and lasts approximately 1 minute and 20 seconds (see Figure 12.1c). The scenario was chosen to compare user experience with scenarios developed ad hoc for the current study to user experience with a professionally developed scenario (Experiment 1). The rollercoaster scenario was also employed in this work to compare the Oculus Rift against two other immersive virtual reality devices.

- iv obstacle scenario: this scenario is a modified version of the train scenario. A set of obstacles were added along the wagon's path. The obstacles, in order of appearance, were (Figure 12.1d): tree branches at eye level, a tree trunk at eye level, a transit barrier on the train tracks, a train crossing on intersecting tracks, a rocky outcrop from a tunnel ceiling at eye level, traffic cones on the train tracks, a wall into which the wagon crashes. The scenario is approximately 2 minutes long. The scenario was developed to assess whether users experienced virtual obstacles as sufficiently looming to require moving out of the way of the obstacle (Kim et al. 2015). I hypothesized that obstacles abruptly appearing at an observer's eye level (tree branches, tree trunk, rocky outcrop) would elicit more movement than obstacles appearing in an observer's lower visual field (transit barrier, traffic cones). I further asked how observers would react to unavoidable virtual impacts (train and wall obstacles).



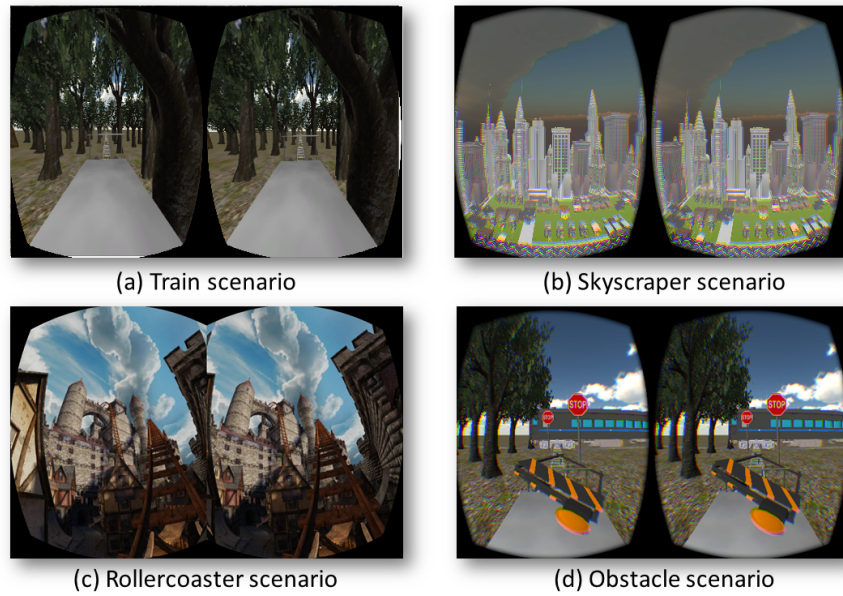


Figure 12.1: **Virtual scenarios.** Example frames from the four virtual scenarios experienced by observers when wearing the Oculus Rift. In Experiment 1, observers were shown the train scenario (a), the skyscraper scenario (b) and the rollercoaster scenario (c). In Experiment 2, observers were shown the obstacle scenario (d).

## Procedures

The perceptual quality of the Oculus Rift was tested in three separate experiments. In Experiment 1 observers were presented with the train scenario, the skyscraper scenario, and the rollercoaster scenario. Observers' heart rate was monitored throughout the duration of the experiment. Following the viewing of all three virtual scenarios, observers were required to rate their experience using a specifically devised questionnaire.

In Experiment 2 observers were presented with the obstacle scenario. Each observer's head movements were monitored throughout the duration of the virtual scenario. Each observer viewed the scenario two consecutive times. Observers were required to fill out the Simulator Sickness Questionnaire (Kennedy et al. 1993) both before (pre-exposure) and after (post-exposure) viewing the virtual scenario.

In Experiment 3 the Oculus Rift was compared against two other low-cost systems for immersive VR: the Google Cardboard and a 47-inch 3DTV. Observers performed three VR sessions on separate days, each session with a different VR system. The order with which observers experienced the three VR systems was counterbalanced across observers. In each session observers were presented with the rollercoaster scenario and sat through three full laps of the rollercoaster ride (approximately 5 minutes). Observers were required to fill out the Simulator Sickness Questionnaire (Kennedy et al. 1993) both before (pre-exposure) and after (post-exposure) each VR session. Observers were also required, at the end of each VR session, to answer four questionnaire items regarding the level of immersivity experienced. The observers' heart rate and head movements were monitored throughout the duration of each

VR session.

## **Experiment 1**

**Heart Rate Measurement** Each subject's heart rate (Gorini et al. 2011) was measured before, during and after the exposure to 3 different virtual scenarios (train scenario, skyscraper scenario and rollercoaster scenario). The heart rate of each observer was manually sampled and recorded using a Intermed SAT-200 pulse oximeter. The device does not digitally store data, hence heart rate measurements were manually transcribed. When observers were viewing the train scenario and the rollercoaster scenario, heart rate was sampled at 5 second intervals. When observers were viewing the skyscraper scenario, heart rate was sampled once when observers were on the ground, once when observers were being lifted, once when observers reached the top of the skyscraper, once when observers were on the ledge of the skyscraper, and once when the observers were in free fall.

**Immersivity Questionnaire** Following the exposure to 3 different virtual scenarios (train scenario, skyscraper scenario and rollercoaster scenario), subjects were asked to compile a self-reported immersivity questionnaire designed for this study. The sensation of presence (or level of immersion) is often measured by means of self-rated questionnaires (Gorini et al. 2011). Several examples can be found in the literature. In the UCL Presence Questionnaire participants are required to provide ratings on a seven-level Likert scale (Slater, Usoh, and Steed 1994). Through the Independent Television Company Sense of Presence Inventory (ITC-SOPI) users are evaluated post-

exposure by providing scores on a five-level Likert scale (Lessiter et al. 2001). The immersivity questionnaire developed for Experiment 1 consisted of 18 questions. The majority of questions required a score on a 1-5 Likert-scale (lower-higher). The 18 questions were as follows:

- Q1** How often do you play videogames? (1-never; 5-very often)
- Q2** How much does the quality of computer graphics affect your enjoyment of virtual reality? (1-not at all; 5-very much)
- Q3** How much do you suffer from fear of heights? (1-not at all; 5-very much)
- Q4** How immersive was the train scenario? (1-not at all; 5-very much)
- Q5** How immersive was the skyscraper scenario? (1-not at all; 5-very much)
- Q6** How strongly did you feel you were truly on the top of a building during the skyscraper scenario? (1-not at all; 5-very strongly)
- Q7** How much vertigo did you experience during the skyscraper scenario? (1-none; 5-very much)
- Q8** When did you experience the most vertigo during the skyscraper scenario? (1-going up; 2-at the top; 3-going down)
- Q9** How immersive was the rollercoaster scenario? (1-not at all; 5-very much)
- Q10** How strongly did you feel you were truly on a rollercoaster ride during the rollercoaster scenario? (1-not at all; 5-very strongly)

- Q11** How much vertigo did you experience during the rollercoaster scenario?  
(1-none; 5-very much)
- Q12** When did you experience the most vertigo during the rollercoaster scenario? (1-going up; 2-going down; 3-during the jump)
- Q13** In which scenario did you experience the most vertigo? (1-train; 2-skyscraper; 3-rollercoaster)
- Q14** How much lag did you notice? (1-none; 5-very much)
- Q15** How much did the screen resolution negatively affect your sense of presence? (1-not at all; 5-very much)
- Q16** Did you notice specific failures of the virtual reality? (1-no; 5-yes)
- Q17** Did you feel more immersed in the virtual environment with time?  
(1-no; 5-yes)
- Q18** How much did knowing you were in a virtual environment reduce your sense of presence? (1-not at all; 5-very much)

## **Experiment 2**

**Head Movement Measurement** Head movements were recorded via the accelerometer integrated within the Oculus Rift while observers were viewing the obstacle scenario for two consecutive sessions. It is well documented that head movements are elicited by visual stimulations (Pélisson et al. 2001). Moreover, many authors in the literature report the use of virtual reality for

sensorimotor training (Adamovich et al. 2009), indicating that visual stimulation in virtual environments elicits observer movements. Several recent studies have addressed the issue ofvection (the visually induced self-motion perception that often emerges as a precursory symptom of motion sickness while viewing moving images in virtual reality environments) through measurement of body sway (Ishio et al. 2015; Sugiura et al. 2015). Head movements were recorded from the Oculus Rift’s built-in accelerometers and sampled at 75 Hz and the time at which each virtual obstacle was presented to the observer was appropriately logged.

**The Simulator Sickness Questionnaire (SSQ)** Subjects were asked to complete the SSQ both before and after being exposed to the two consecutive presentations of the obstacle scenario. The SSQ was introduced by Kennedy et al. (1993) and it is commonly used to assess the discomfort induced by the virtual simulation. Participants give a score from 0 to 3 (in ascending order "None", "Slight", "Moderate" and "Severe") to a set of symptoms before and after virtual experience. Every symptom belongs to one or more of 3 classes of discomfort: nausea, oculomotor problems, and disorientation. By combining scores from multiple symptoms a partial score can be computed for each class of discomfort. These 3 partial scores can then be combined to produce a total SSQ score. The SSQ is a widely applied measurement tool in research studying simulator sickness and cybersickness (for a recent review see Davis, Nesbitt, and Nalivaiko 2014)

### **Experiment 3**

**Heart Rate Measurement** The HS-2BT heart rate monitor samples the user's heart rate every 5 seconds and logs a time stamp for each measure. Observers' heart rate was recorded for the full duration of each VR session. Observer's baseline heart rate was computed by acquiring samples from approximately 30 seconds before the start of each VR session.

**Head Movement Measurement** Because Experiment 3 was meant to compare different VR devices, the Oculus Rift internal accelerometers could not be employed to measure head movements. Instead, the Microsoft Kinect (version 1) was employed to compute and track observers' head position during each VR session. The Microsoft Kinect SDK provides 3D skeletal joint tracking at 30 Hz. Throughout each VR session the 3D position data of the skeletal joint corresponding to the observer's head was recorded. Each observer's baseline head movement was computed by acquiring samples from approximately 30 seconds before the start of each VR session.

Heart rate and head movement data were appropriately synchronized with each VR session by an ad hoc software routine which logged the start time of each VR session.

**The Simulator Sickness Questionnaire (SSQ)** As in Experiment 2, subjects were asked to complete the SSQ both before and after being exposed to the three consecutive presentations of the rollercoaster scenario.

**Immersivity Questionnaire** As in Experiment 1, following each VR session subjects were asked to complete a self-reported immersivity question-

naire. This immersivity questionnaire was different from the one used in Experiment 1. It consisted of 4 questions which required a score on a 1-5 Likert-scale (lower-higher). The 4 questions are as follows:

**Q1** How much do you suffer from fear of heights? (1-not at all; 5-very much)

**Q2** How much vertigo did you experience during the virtual environment?  
(1-not at all; 5-very much)

**Q3** How immersive did you find the virtual environment? (1-not at all;  
5-very much)

**Q4** Did you feel more immersed in the virtual environment with time? (1-  
not at all; 5-very much)

## Subjects

Twenty-six volunteers, with ages ranging between 20 and 40 years, were recruited to participate in Experiment 1. Twelve different volunteers, with ages ranging between 20 and 40 years, were recruited to participate in Experiment 2. Prior to Experiment 3 I performed sample size calculations and determined that a sample size of eighteen subjects would be sufficient to detect differences of the same magnitude as those observed in the data from Experiments 1 and 2 at the 95% confidence level with 85% power. Thus, eighteen additional volunteers, with ages ranging between 23 and 63 years, were recruited to participate in Experiment 3. All participants had normal or corrected to normal vision and reported normal stereo-vision.



## 12.3 Results

### Experiment 1

**Heart Rate** Resting state heart rate (before and after exposure to the virtual scenarios) for each subject was compared to the heart rate recorded during each scenario. Figure 12.2 shows heart rate at each exposure condition (baseline pre-exposure, train scenario, skyscraper scenario, rollercoaster scenario, baseline post-exposure) averaged across observers. Data were analyzed with a one way, within-subjects ANOVA, which showed a significant effect of exposure condition ( $F(4,100)=27.6$ ,  $p<10^{-14}$ ). Post-hoc analysis was performed with a Tukey-Kramer single-step, multiple comparison procedure. Average heart rate prior to viewing any of the virtual scenarios was 83.3 beats/min (95% CI [81.2 85.3] beats/min), while the average heart rate at rest after viewing the virtual scenarios was 77.6 beats/min (95% CI [75.2 77.5] beats/min); the difference between these values was statistically significant ( $p<0.05$ ). This is possibly due to the fact that subjects were slightly excited before starting the experiment (Slater et al. 2006; Young, Adelstein, and Ellis 2007). Critically, heart rate was clearly greater when observers were viewing virtual scenarios compared to when observers were in a resting state ( $p<10^{-5}$  for all comparisons between heart rate during the virtual scenarios and baseline heart rate post-exposure).

**Immersivity Questionnaire** The questionnaire data collected was investigated by inspecting the response distribution for each question. Figure 12.3 shows the response distribution histograms for all 18 questionnaire questions.

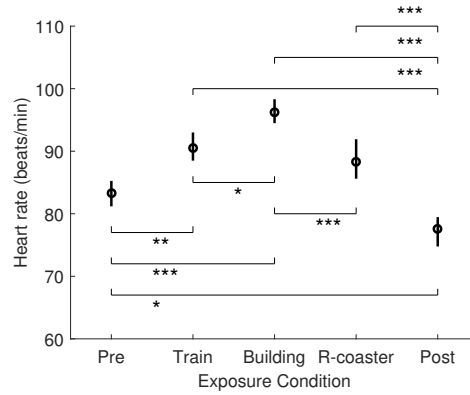


Figure 12.2: **Heart rate data from Experiment 1.** Heart rate for each condition (baseline pre-exposure, train scenario, skyscraper scenario, rollercoaster scenario, baseline post-exposure) in Experiment 1, averaged across observers (n=26). Error bars are 95% bootstrapped within-subject confidence intervals. (\*  $p<0.05$ ; \*\*  $p<0.01$ ; \*\*\* $p<0.001$ ).

Most of the subjects found the skyscraper and rollercoaster scenarios to be realistic and immersive (questions Q5, Q6, Q9 and Q10 all have mode 4). Lag was rarely reported (Q14). The resolution of the virtual scenarios was somewhat of an issue (Q15). Approximately half of the subjects noticed glitches in the virtual reality rendering (Q16). The majority of subjects reported a greater feeling of immersion as they spent more time in the virtual environment (Q17). The distribution of responses regarding the experience of vertigo during the exposure to virtual scenarios involving heights was fairly uniform (questions Q7 and Q11).

I further correlated (using Spearman's correlation) observers' self-reported fear of height with their experience of vertigo during the skyscraper and rollercoaster scenarios (Figure 12.4). This was done to investigate the effec-

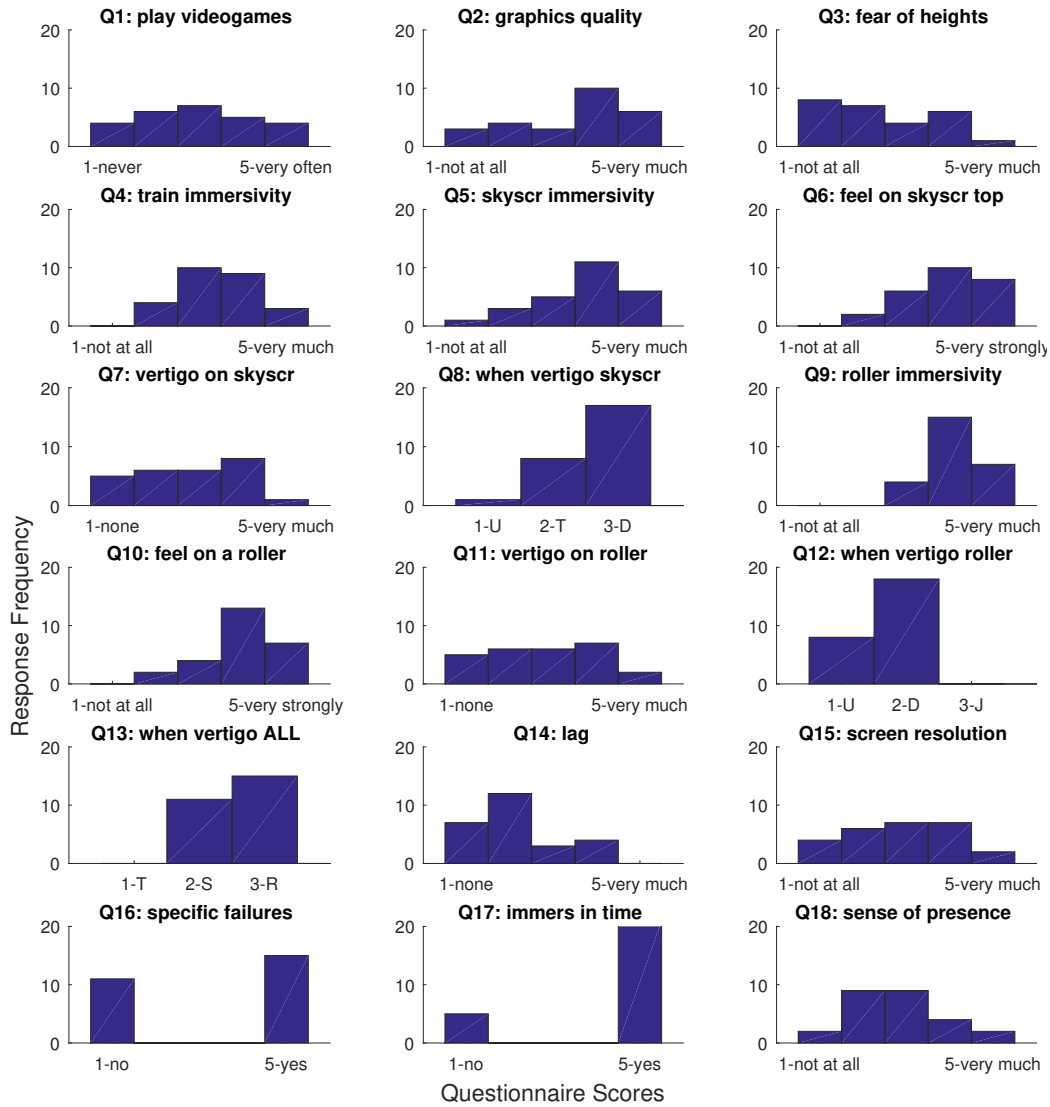


Figure 12.3: **Immersivity Questionnaire data from Experiment 1.** Response distribution histograms for all 18 questionnaire items administered to observers (n=26) in Experiment 1.

tiveness of virtual scenarios experienced with the Oculus with respect to triggering phobias. There was no significant correlation between observers' fear of heights and vertigo experienced during the skyscraper scenario ( $r=0.13$ ,  $p=0.52$ ). Conversely, there was a strong relationship between observers' fear of heights and vertigo experienced during the rollercoaster scenario ( $r=0.74$ ,  $p<10^{-5}$ ).

The rollercoaster scenario, which was professionally developed, elicited a greater level of immersion than the other scenarios (cf. histogram for Q9 with histograms for Q4 and Q5 in Figure 12.3). It is thus sensible that the greater level of immersion triggered a stronger reaction in those observers already susceptible to the fear of heights.

In the Skyscraper scenario, abruptly dropping observers might have gone beyond inducing vertigo by triggering instead shock. Thus, we verified whether our observers underwent a greater physiological response to falling than to simply being atop the skyscraper. We compared heart-rate measurements taken when observers were on the ledge of the skyscraper to heart-rate measurements taken when observers were in free fall. We found no difference in heart rate between when observers were on the ledge of the skyscraper compared to when observers were in free fall ( $p=0.71$ , paired samples t-test).

### **Relationship between Heart Rate and Immersivity Questionnaire**

A multiple linear regression was calculated to assess the relationship between Heart Rate Increment (dv) and Questionnaire Scores (predictors). Results of the model fitting are reported in Table 12.1. A significant regression equation was found ( $F(18,6)=9.59$ ,  $p<0.01$ ), with an  $R^2$  of 0.9664. Questionnaire

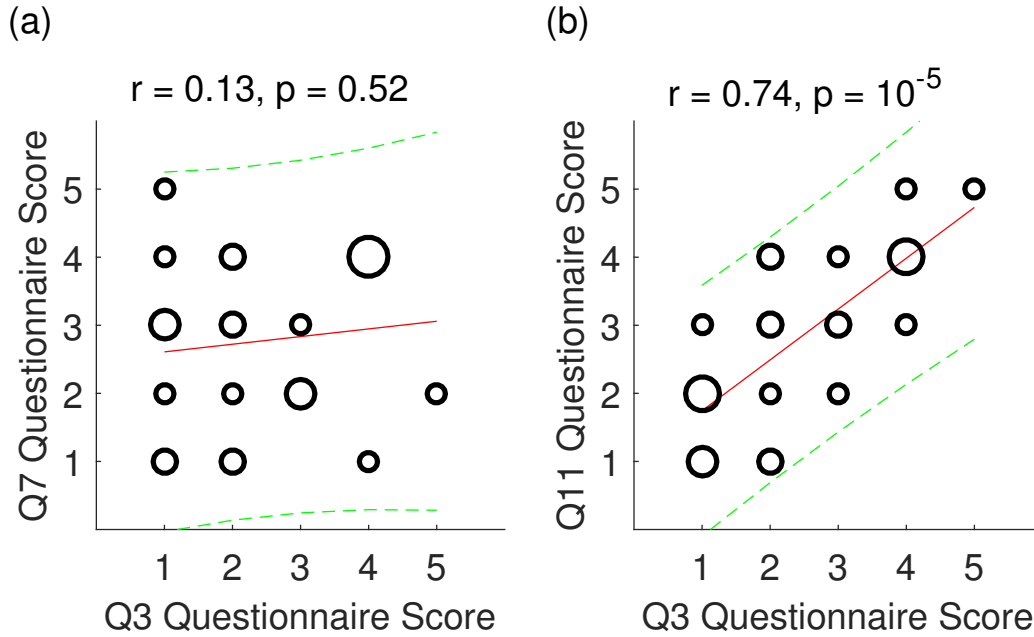


Figure 12.4: **Correlation between observers' fear of height and vertigo in Experiment 1.** Correlation between observers' self-reported fear of height and their experience of vertigo during the (a) skyscraper scenario and (b) rollercoaster scenario. Black circles are vertigo scores for the (Q7) skyscraper and (Q11) rollercoaster scenarios plotted against the (Q3) fear of heights scores for all subjects ( $n=26$ ) in Experiment 1. The size of each circle represents the number of occurrences for each data point. Red lines are best fitting linear regression lines, bounded by 95% confidence intervals of the fit (green dashed lines).

questions that were found to be significant predictors of heart rate increment after false discovery rate (FDR) adjustment (Benjamini and Hochberg 1995) are highlighted in red in Table 12.1.

This analysis confirms that self-reported measurements of the quality of the virtual reality are related to the quantitative assessment of the physiological response to virtual environments. The two metrics, questionnaire responses and heart rate increase, taken together suggest that observers had an immersive and physically stimulating virtual reality experience.

## Experiment 2

**Head Movement Measurement** While observers viewed the obstacle scenario, head movements were recorded using the Oculus Rift’s built-in accelerometer. The accelerometer supplies the acceleration of the device (and thus of the user’s head) along the medio-lateral (X), antero-posterior (Z) and longitudinal (Y) axes. The modulus of the acceleration value was thus employed as a single metric of the instantaneous amount of head movement.

Figure 5 (top) shows the modulus of the head acceleration throughout a whole viewing of the obstacle scenario for a representative subject. The time points at which the obstacles would virtually impact the observer or the wagon are shown as vertical dotted lines. A threshold value (horizontal dotted line) was set equal to mean acceleration plus twice the standard deviation: the instantaneous accelerations that exceed this threshold are highlighted as green dots.

Qualitatively we note that in temporal correspondence with the appearance of the obstacles (a few seconds prior to the virtual impact between the

	Estimate	Standard Error	<i>t</i> statistic	p-Value
<b>Intercept</b>	4.5	12.9	0.3	0.782
<b>Q1</b> play videogames	-3.8	1.1	-3.5	0.030
<b>Q2</b> graphics quality	-2.3	0.9	-2.4	0.102
<b>Q3</b> fear of heights	-1	1.9	-0.5	0.738
<b>Q4</b> train immersivity	1.9	1.3	1.5	0.246
<b>Q5</b> skyscr immersivity	6.5	1.6	4.2	0.024
<b>Q6</b> feel on skyscr top	2.6	1.5	1.8	0.189
<b>Q7</b> vertigo on skyscr	6.6	1.2	5.3	0.017
<b>Q8</b> when vertigo skyscr	-12.5	3	-4.1	0.024
<b>Q9</b> roller immersivity	0.4	2.1	0.2	0.861
<b>Q10</b> feel on a roller	-3.5	1.8	-1.9	0.179
<b>Q11</b> vertigo on roller	1	1.2	0.9	0.533
<b>Q12</b> when vertigo roller	-9.4	3.9	-2.4	0.102
<b>Q13</b> when vertigo ALL	19.6	3.2	6.2	0.016
<b>Q14</b> Lag	-5.1	1.4	-3.7	0.028
<b>Q15</b> screen resolution	3.4	1.9	1.8	0.189
<b>Q16</b> specific failures	-8.9	2.3	-3.8	0.027
<b>Q17</b> immers in time	-16.9	3.8	-4.4	0.024
<b>Q18</b> sense of presence	-0.7	1.8	-0.4	0.782

Table 12.1: **Multiple Linear Regression assessing the relationship between Heart Rate and Questionnaire Scores.** For each term included in the linear model (rows), the table reports the associated coefficient Estimate, the Standard error of the coefficients, the *t* statistic for each coefficient estimate, and the corresponding FDR adjusted p-Value. P-Values in red correspond to Questionnaire Scores found to be significant predictors of Heart Rate Increment.

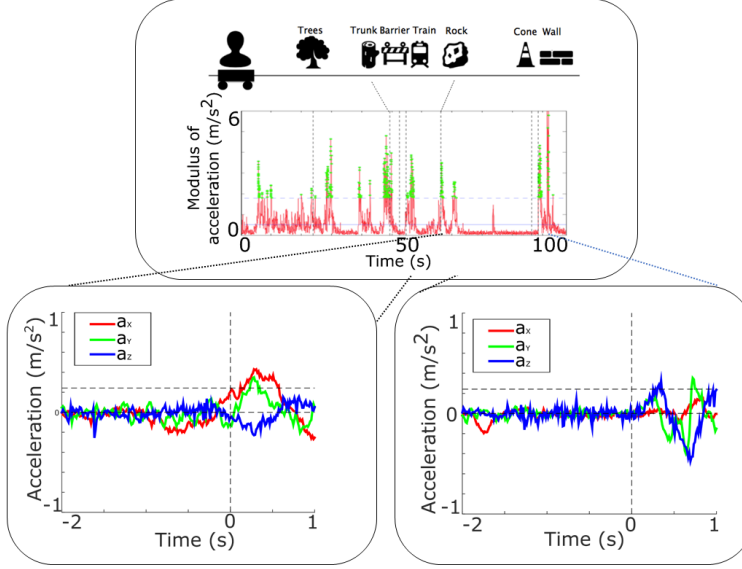


Figure 12.5: **Accelerometer data from Experiment 2.** (top) Modulus of head acceleration as a function of time from scenario onset (red curve) plotted for one representative observer during one session of Experiment 2. Horizontal blue line represents mean acceleration throughout the session. Peak accelerations, defined as acceleration values greater than the mean plus twice the standard deviation (blue dashed horizontal threshold line) are highlighted in green. Vertical dashed lines denote the times at which each obstacle appeared. (bottom) Head acceleration as a function of time from obstacle onset for each axis separately for (left) the rock obstacle and (right) the wall obstacle. Red, green and blue curves represent acceleration along X, Y, and Z axes, respectively, averaged across 12 observers participating in Experiment 2. Horizontal dashed lines highlight null level of acceleration and the average baseline level of acceleration. Vertical dashed line highlights the time at which the obstacles reach the smallest distance from the observer.



obstacles and the observer) there is an increase in the magnitude of head movements with respect to the average magnitude of movement, and that not all events cause the same intensity of head movements in the observer.

To verify the influence of each obstacle on the observer's head movements we selected the acceleration signal from two seconds prior to one second after the virtual impact with each obstacle. We compared these segments of the signal with the average amount of acceleration measured in all segments of the scenario not containing obstacles. To exemplify, the two insets of Figure 5 present the head acceleration for one representative subject measured along all three axes in correspondence of two obstacles: the rocky outcrop and the wall. The red, green, and blue curves show the average acceleration along the X, Y, and Z axes respectively. The traces in the inset on the left show that observers moved along the X axis, presumably to avoid the rocky outcrop suddenly appearing from the side. The inset on the right highlights an oscillatory movement along the Z axis after observers have virtually struck the wall obstacle in a frontal collision.

Figure 12.6 shows the mean of the acceleration averaged across all the subjects and all the events with respect to the baseline (dashed horizontal line). On average, there is a small increase in the amount of motion when the obstacles appear (prior to the virtual collisions), and a sustained increase in the amount of motion during and following the virtual collisions.

To statistically verify whether different obstacles induced different amount of motion in the observers, for each obstacle we computed the average acceleration from two seconds prior to each obstacle to one second after each obstacle. This value was averaged across conditions for every observer, and

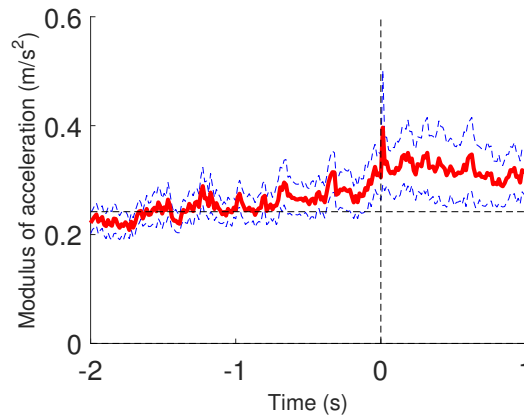


Figure 12.6: **Head movements as a function of time from obstacle onset.** Modulus of head acceleration as a function of time from obstacle onset averaged across all obstacles. Red curve is the average across sessions and observers participating in Experiment 2 (n=12) bounded by  $\pm 1$  Standard Error (dashed blue line). Horizontal dashed line is the average baseline level of acceleration. Vertical dashed line highlights the time at which the obstacle virtually impacts the observer.

then across observers (Figure 12.7). These data were then analyzed with a one-way, within-subjects ANOVA, which showed a significant effect of obstacle type ( $F(6,66)=5.0$ ,  $p<0.001$ ). Post-hoc analysis was performed with a Tukey-Kramer single-step, multiple comparison procedure. The Trees obstacle elicited a greater reaction than both the Barrier obstacle ( $p<0.05$ ) and the Cone obstacle ( $p<0.01$ ). The Cone obstacle, also elicited a smaller reaction than Trunk obstacle ( $p<0.05$ ) and the Rock obstacle ( $p<0.05$ ).

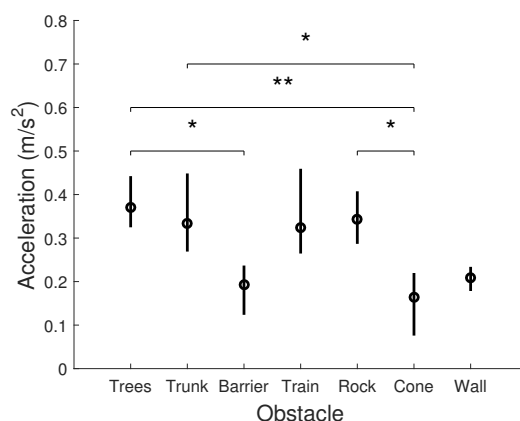


Figure 12.7: **Head movements for virtual obstacles.** Average head acceleration in correspondence of all 7 virtual obstacles. The data are the mean for 12 observers. Error bars are 95% bootstrapped within-subject confidence intervals. (\*  $p<0.05$ ; \*\*  $p<0.01$ ).

**Simulator Sickness Questionnaire** VR environments may cause users to experience sickness, visual fatigue, and disorientation. To evaluate the Oculus Rift regarding these issues in our scenarios, we asked subjects to fill out the Simulator sickness questionnaire (SSQ), both before and after the exposure to the virtual environment. Participants rated how strongly

they were experiencing 16 different symptoms on a scale from none to slight, moderate or severe.

The SSQ provides a total severity score (TS) as well as partial scores for three subscales, nausea-related subscore (N), oculomotor-related subscore (O), disorientation-related subscore (D). The TS is obtained by combining the scores from the three subscales, and provides an index of the overall symptoms of cybersickness experienced by the users of virtual reality systems. The three subscales provide diagnostic information about particular symptom categories: the nausea subscale includes symptoms such as increased salivation, sweating, nausea, stomach awareness, and burping; the oculomotor subscale includes fatigue, headache, eyestrain, and difficulty focusing; the disorientation includes vertigo, dizziness (with eyes open as well as with eyes closed), and blurred vision.

The subscores provide differential information about participants' experience of symptoms and are useful for determining the particular pattern of discomfort produced by a given simulator, or virtual reality environment. All scores have as their lowest level a natural zero (no symptoms) and increase with increasing symptoms reported.

The three subscores (N, O, D) as well as the TS, averaged across observers have been computed for the pre-exposure and the post-exposure conditions. The results are summarized in Table 12.2.

Subscores, as well as TS, are slightly smaller in the post-exposure condition than in the pre-exposure condition, and this difference is not statistically significant (TS pre-exposure vs TS post-exposure,  $p=0.28$ , paired samples t-test). Thus observers did not experience any measurable simulator sickness

	Nausea (N)	Oculomotor (O)	Disorientation (D)	Total Score (TS)
<b>Pre-exposure</b>	31.8 [25, 37.5]	43.6 [37.5, 50.8]	36.0 [28.4 47]	43.6 [28.3, 42.6]
<b>Post-exposure</b>	23.4 [18.1, 31.2]	37.2 [29.7, 44.1]	27.8 [15.7 36]	35.0 [28.3 42.5]

Table 12.2: **Simulator Sickness Questionnaire scores from Experiment 2.** Simulator Sickness Questionnaire subscores and Total Score (columns) computed for the pre-exposure and post-exposure conditions (rows) in Experiment 2. Data are the mean across all observers. Values in brackets are 95% bootstrapped within-subject confidence intervals.

after the exposure to VR experienced through the Oculus Rift. It is worth noting that the maximum total score achievable on the SSQ is approximately 300. The scores observed during our experiments are low with respect to this maximum value (less than 15%), in line with baseline scores reported in the literature for subjects not affected by sickness symptoms (Solimini 2013).

### Experiment 3

**Heart Rate** Figure 12.8 shows heart rate as a function of time from the start of the VR session averaged across observers for each VR system tested. Qualitatively we can observe that with all three systems there appeared to be an initial increase in heart rate as soon as the virtual scenario commenced. When observers were wearing the Oculus Rift (blue trace), on average heart rate steadily increased throughout the VR session. When observers were watching the 3DTV (red trace), after the initial increase heart rate stabilized on an elevated level with respect to baseline. When observers were

wearing the Google Cardboard system instead (green trace), following the initial increase in heart rate there appeared to be, on average, a gradual decrease.

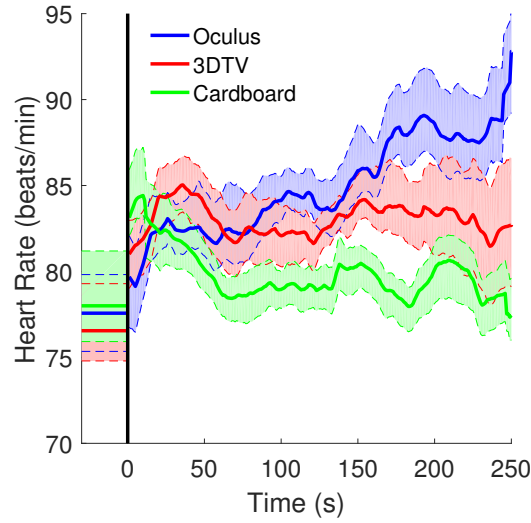


Figure 12.8: **Heart rate in Experiment 3.** Heart rate as a function of time from the start of to the virtual rollercoaster ride in Experiment 3. Blue, red and green curves are heart rate measurements averaged over all observers taken throughout each experimental session with the Oculus Rift, the 3DTV and the Google cardboard systems respectively. Shaded regions represent 68% bootstrapped within-subject confidence intervals. Vertical black line highlights the start of the VR session. Traces prior to time zero are baseline heart rate computed from heart rate measurements taken before the start of each VR session.

To statistically investigate these data I computed, for each observer and each session, the baseline heart rate as the average heart rate prior to the VR session, and the test heart rate as the average heart rate during the whole

VR session. These data were then analyzed with a 2x3 (Session Timeperiod x VR System), within-subjects ANOVA, which showed a significant main effect of Session Timeperiod ( $F(1,34)=5.6$ ,  $p<0.05$ ), no main effect of VR System ( $F(2,34)=0.5$ ,  $p=0.6$ ), and no significant interaction between Session Timeperiod and VR System ( $F(2,34)=0.7$ ,  $p=0.5$ ). This analysis indicates that with all three VR systems heart rate was significantly greater during the VR session than at baseline. This heart rate increase due to the VR experience was however not significantly different between VR systems.

**Head Movement Measurement** Prior to and during each VR session in Experiment 3 observers' head position was tracked using a Microsoft Kinect sensor. The skeletal joint tracking provided by the Microsoft Kinect SDK was employed to record the 3D position of the skeletal joint corresponding to an observer's head. These data were used to compute the average amount of head movement prior to and during the VR session. Baseline head movement was computed on the data from the 30 seconds prior to the VR session, when observers were instructed to quietly sit with their eyes closed and wait for the VR session to commence. Test head movement was computed on the data acquired during the whole VR session. A metric of head movement during the baseline and test periods was computed as follows. First, the variance of the 3D head position was computed along the three axes separately. Then, the variances were summed and the square root of the summed variances was taken as the final metric.

Figure 12.9 shows baseline and test head movement averaged across observers for each VR system tested. These data were analyzed with a 2x3

(Session Timeperiod x VR System), within-subjects ANOVA, which showed a significant main effect of Session Timeperiod ( $F(1,34)=14.2$ ,  $p<0.001$ ), no main effect of VR System ( $F(2,34)=3.25$ ,  $p=0.05$ ), and a significant interaction between Session Timeperiod and VR System ( $F(2,34)=4.07$ ,  $p<0.05$ ). Post-hoc analysis was performed with a Tukey-Kramer single-step, multiple comparison procedure. As can be clearly seen in Figure 9, both the Oculus Rift and the Google Cardboard systems produced a significant increase of head movement from baseline ( $p<0.05$  and  $p<0.01$  respectively), whereas head movement was not significantly different between the baseline and test periods when observers were viewing the 3DTV.

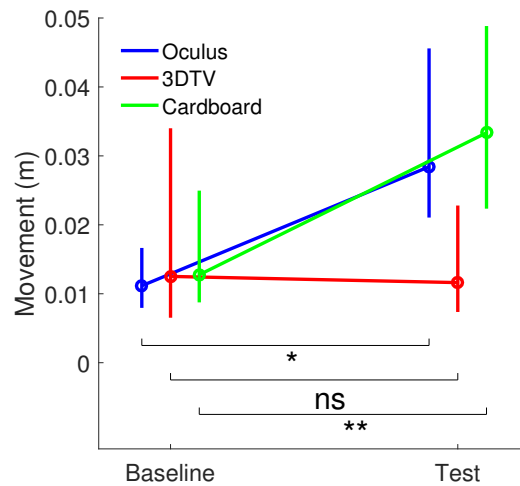


Figure 12.9: **Head movements in Experiment 3.** Head movement prior to (baseline) and during (test) the virtual rollercoaster ride in Experiment 3. Data are the mean across all observers for the Oculus Rift (blue), 3DTV (red) and Google Cardboard (green) VR session. Error bars are 95% bootstrapped within-subject confidence intervals. (\*  $p<0.05$ ; \*\*  $p<0.01$ ).



**Immersivity Questionnaire** Following each of the three VR sessions observers were asked to respond to the Immersivity questionnaire items designed for Experiment 3. The questionnaire data collected was investigated by inspecting the response distribution for each question and by comparing the response across VR systems via Bonferroni corrected paired t-tests (planned comparisons). Figure 12.10 shows the response distribution histograms for the questions regarding the VR experience. The insets of each subfigure show the results of the planned comparisons.

Observers experienced some vertigo with the Oculus Rift system (Figure 12.10a, blue histogram exhibits mode 3) and this was significantly more than what observers experienced with either the 3DTV (red,  $p < 10^{-4}$ ) or the Google Cardboard system (green,  $p < 10^{-4}$ ). The 3DTV and the Google Cardboard system elicited similarly little or no vertigo ( $p = 0.04$ ). Observers found the VR experienced with the Oculus Rift to be immersive (Figure 12.10b, blue histogram exhibits mode 4). Furthermore, VR experienced with the Oculus Rift was significantly more immersive than with either the 3DTV ( $p < 10^{-5}$ ) or the Google Cardboard system ( $p < 10^{-6}$ ). The 3DTV and Google Cardboard system were reported to be equally as immersive ( $p = 0.09$ ). Observers felt similarly more immersed in the VR scenario with time (Figure 12.10c) with both the Oculus Rift and the 3DTV ( $p = 0.8$ ). Observers did not report feeling more immersed in the VR scenario with time with the Google Cardboard system, and this reported feeling of immersion was smaller than that reported with either the Oculus Rift ( $p < 0.01$ ) or the 3DTV ( $p < 0.01$ ).

To summarize, observers felt more vertigo and more immersion with the Oculus Rift than with either the 3DTV or the Google Cardboard. Further-

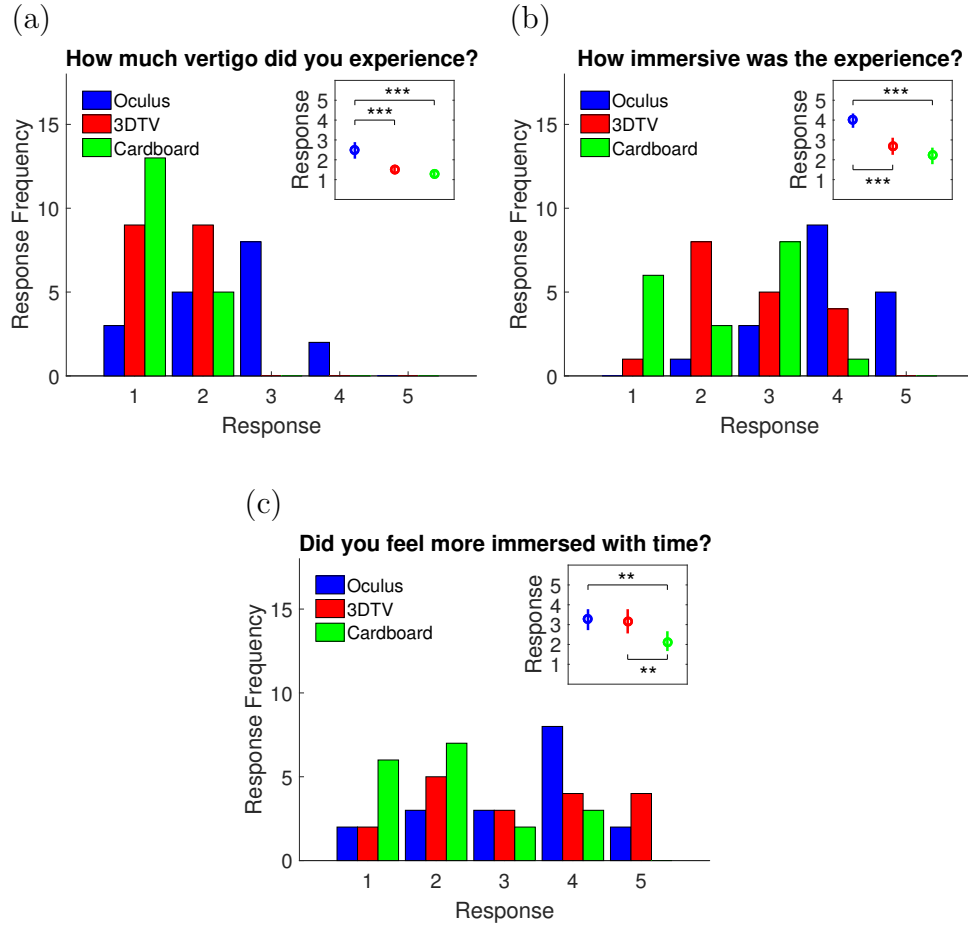


Figure 12.10: **Immersivity questionnaire in Experiment 3.** Response distribution histograms for the three (a-c) immersivity questionnaire items administered to observers ( $n=18$ ) in Experiment 3. Blue, red and green bars are the response distributions regarding the Oculus Rift, the 3DTV and the Google Cardboard respectively. The insets of each figure present the same data as mean response across observers bounded by 95% bootstrapped within-subject confidence intervals in order to highlight statistically significant comparisons (\*\*  $p<0.01$ ; \*\*\*  $p<0.001$ ).

more, observers felt somewhat more immersed with time with the Oculus Rift and the 3DTV, less so with the Google Cardboard.

I further correlated (using Spearman's correlation) observers' self-reported fear of height with their experience of vertigo during each VR session with all three VR systems. There was no relationship with any of the three VR systems (Oculus Rift:  $r=0.21$ ,  $p=0.41$ ; 3DTV:  $r=0.23$ ,  $p=0.35$ ; Google Cardboard:  $r=0.43$ ,  $p=0.07$ ). However, after removing data from three outliers who reported not suffering from fear of height but experiencing strong vertigo during the Oculus Rift VR session, I found a strong relationship between observers' fear of heights and vertigo experienced during the rollercoaster scenario experienced with the Oculus Rift ( $r_s=0.69$ ,  $p<0.01$ ). This observer correlation is in line with the results from Experiment 1.

**Simulator Sickness Questionnaire** The SSQ was administered to all subjects prior to and following each VR session. Figure 12.11 presents the SSQ Total Score averaged across observers for each VR system and for the pre-exposure and the post-exposure conditions. Solely the Total Score is reported as we found the pattern of results to be the same when broken down into the three subscores. Qualitatively, Figure 12.11 shows that on average post-exposure TS was higher than the pre-exposure TS for all VR systems, and that the Oculus Rift had a higher average post-exposure SSQ score than either the 3DTV or the Google Cardboard.

These data were analyzed with a 2x3 (Exposure condition x VR System), within-subjects ANOVA, which showed a significant main effect of Exposure condition ( $F(1,34)=9.8$ ,  $p<0.001$ ), no main effect of VR System

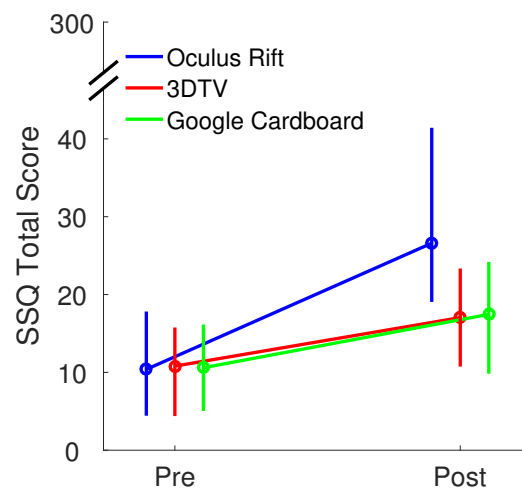


Figure 12.11: **Simulator Sickness Questionnaire scores in Experiment 3.** Simulator Sickness Questionnaire scores Pre and Post exposure to the VR session in Experiment 3. Data are the mean SSQ TS across all observers for the Oculus Rift (blue), the 3DTV (red) and the Google Cardboard (green) VR sessions. Y-axis break highlights the full scale of the SSQ TS.

( $F(2,34)=0.9$ ,  $p=0.4$ ), and no significant interaction between Exposure condition and VR System ( $F(2,34)=1.1$ ,  $p=0.4$ ). This analysis indicates that exposure to the virtual rollercoaster ride induced some Simulator Sickness regardless of the VR system employed. It is worth noting however that the average post-exposure scores are low and within the range of baseline scores reported in the literature for subjects not affected by sickness symptoms (Solimini 2013). As a final note, although on average observers reported little or no Simulator Sickness symptoms, one observer in the Oculus Rift VR session had to terminate the virtual rollercoaster ride after two laps because of a severe bout of nausea that disappeared after a few minutes.

## 12.4 Discussion

In this chapter I have investigated the perceptual quality of the Oculus Rift head-mounted-display. Specifically, I aimed to characterize (i) the level of Immersivity provided by the device and (ii) the potential Cybersickness induced by the device. I evaluated whether the Oculus Rift is able to make the user feel like he/she is in a real world scenario, and if it is possible to elicit in the users the sensation of presence via the virtual stimuli rendered by the device. Furthermore, I assessed whether VR experienced through the Oculus Rift induces physical discomfort in the user. Finally, I compared the levels of Immersivity and Cybersickness induced by the Oculus to those induced by two other low-cost VR systems, a widescreen 3DTV and a Google Cardboard.

The investigation has provided the following principal results.

The heart rate of human observers increased during the exposure to virtual scenarios experienced via the Oculus Rift. This result is consistent with previous literature that links physiological activation with levels of immersion (Gorini et al. 2011).

The self-reported answers to a specifically devised immersivity questionnaire show that the majority of participants felt the experience was immersive and realistic.

A significant correlation was observed between self-reported fear of heights and the sensation of vertigo experienced in one of two virtual scenarios involving heights. The fact that a significant relationship was observed only with a professionally developed virtual scenario sensibly suggests that the level of detail with which virtual worlds are created, and not solely of the employed HMD technology, may play a significant role in the level of immersion experienced by users.

A multiple linear regression highlighted a relationship between Heart Rate Increment (dv) and Questionnaire Scores (predictors). This result further validates the use of physiological activation monitoring to assess levels of immersion. Both metrics taken together suggest that observers had a reportedly immersive and physically stimulating virtual reality experience.

Observers were reactive to virtual objects placed in their path. Different kinds of obstacles elicited different reactions. Specifically, virtual obstacles suddenly presented in front of the observers made them attempt to avoid them, consistent with what happens in a real world situations. This is further indication that the Virtual Reality experienced through the Oculus Rift is sufficiently realistic and immersive to fool users into believing they might be

struck by virtual obstacles.

The Oculus Rift did not induce simulator sickness symptoms when observers were viewing the train scenario. I employed a widely adopted test, i.e. the Simulator Sickness Questionnaire, and found that no appreciable sickness symptoms were induced by the Oculus Rift when the exposure to virtual stimuli was short and did not induce excessive amounts of vection.

Compared to other VR systems, namely a widescreen 3DTV and a Google Cardboard, the Oculus Rift elicited a greater sensation of immersion and similar levels of physiological activation. The rollercoaster ride in Experiment 3 likely induced a stronger sensation of vection than the train scenario in Experiment 2, and accordingly all three VR systems in Experiment 3 produced mild and similarly low levels of simulator sickness symptoms. On average, the Oculus Rift elicited higher simulator sickness scores than the other VR systems. Even though this difference was not statistically significant, it is nonetheless consistent with the notion that the Oculus induced a stronger sensation of immersivity and vection compared to the other systems.

The preliminary investigation of the Oculus Rift HMD described in this chapter suggests that the Oculus has great potential for employment in an array of basic research (e.g. Kim et al. 2015) and clinical applications (e.g. H. G. Hoffman et al. 2014). Nevertheless, multiple aspects regarding the Oculus need to be further investigated. The different user reactions observed to the different virtual scenarios employed suggest that the detail and setting of the virtual worlds play a role in user experience with HMD. More detailed investigations of these factors are warranted. The analyses presented here validate the assumed link between physiological activation and user experi-

ence. However, it would be valuable to examine the relationships between subjective user experience and additional objective physiological measurements such as continuous heart rate measurements, electrodermal activity and event-related electrocardiography (Phillips et al. 2012; Repetto and Riva 2011).

In the Skyscraper scenario, abruptly dropping observers could have gone beyond inducing vertigo by triggering instead a fight-or-flight type stress response. There was no evidence of this, although there are informal reports of shockingly vivid experiences with head mounted displays, where users appear to completely forget where they are and seem truly convinced of being in life-threatening situations. It might thus be of interest to investigate in future studies which interactions between the content of the VR stimuli and the VR technologies may induce fight-or-flight type stress responses in users.

The rollercoaster scenario, with virtual stimuli that inducedvection, induced low yet detectable amounts of simulator sickness, which were greatest when users viewed the virtual content with the Oculus Rift. The increase in simulator sickness may be the consequence of increasing the veracity of VR simulation in head mounted display systems, which in turn leads to an increase in visual and vestibular conflict in users. It will thus be important to continue to monitor and investigate these issues as VR technology further develops.

Finally, this investigation focused on short ( $< 5$  minute-long) durations of exposure to Virtual Reality, and previous research has shown that using bi-ocular HMD devices (which do not contain stereoscopic depth) for up to 30 minutes produces no adverse symptoms (Rushton, Mon-Williams, and Wann



1994). Short exposure durations are appropriate for designing basic research studies and clinical applications. However, it has been shown that undesired side-effects appear with long exposure to the virtual reality (Steinicke and Bruder 2014). These results therefore suggest that shorter experimental or therapeutic sessions may be useful to avoid simulator sickness. Furthermore, some studies indicate that reports of motion sickness after immersion in a virtual environment are much greater when both pre and post questionnaires are given than when only a posttest questionnaire is used (Young, Adelstein, and Ellis 2007). Thus, a more in depth investigation of the link between exposure duration, experimental design, and simulator sickness, with the Oculus as well as with other VR technology, is warranted.

## **Part 7: Final Conclusions and Future Work**

## Chapter 13

# Summary of Results and Conclusions

The goal of the work presented in this thesis was to further our understanding of oculomotor control and visual perception in three dimensional space, particularly in the context of common visual deficits such as myopia, amblyopia and strabismus. In Part 1 (Chapters 3 and 4), I reviewed the usefulness of simulated dioptric blur for depth perception and binocular fusion of rendered disparity depth content. In Part 2 (Chapters 5, 6 and 7), I discussed the importance of peripheral blur and depth information with regards to the development of myopia. In Part 3 (Chapters 8, 9 and 10), I investigated oculomotor recalibration, hand-eye coordination and the neural processing underlying optic flow computations. Lastly, in Part 4 (Chapters 11 and 12) I evaluated two novel commercial devices that have the potential of translating basic oculomotor and vision research into clinical and commercial applications.

In Chapter 3 I have shown that simulated blur does not facilitate, and actually hinders, the perception of depth in image patches of naturalistic stimuli. Thus it is unlikely that rendered generic defocus blur will serve to enhance the fine perception of relative depth in 3D virtual reality technology. Rendered blur might however play other roles in virtual reality rendering techniques, as previous research has shown that blur gradients can modulate the perception of egocentric distance, an example of this is the tilt-shift miniaturization effect (Vishwanath and Blaser 2010). In Chapter 4 I have shown that gaze-contingent dioptric blur modifies eye-movement behavior and facilitates binocular fusion. The system I have described simulates the patterns of blur in the visual periphery due to changes in accommodation. The findings reported in Chapter 4 thus suggest that interruptions of the natural pattern of peripheral retinal blur may adversely affect binocular fusion.

The distributions of peripheral retinal blur may also play a role in the development of myopia. In Chapter 5 I have discussed how it is established that spending time outdoors is protective against myopia (Jones et al. 2007; Rose et al. 2008; Dirani et al. 2009), and I have shown that there are substantial differences in the patterns of peripheral blur arising from outdoors and indoors environments. I have further shown that elongated myopic eyes have greater amounts of peripheral blur, and that this may lead to measurable deficits in peripheral contrast sensitivity of myopes compared to emmetropes. In Chapter 6 I have then shown that myopes have, on average, worse sensitivity to blur than emmetropes monocularly. This impairment might be due to neural adaptation to greater levels of retinal blur in my-

opes, or it might play a causative role in the development of myopia. If the eyes of developing myopes are less sensitive to blur, this might lead to errors in the processes regulating eye growth. Furthermore, there is likely to be an interplay between blur sensitivity and accommodation in myopia since in Chapter 7 I have found that accommodation is weakened and less stable in myopes when viewing simulated disparity. Unstable accommodation and slower vergence eye movements in myopes viewing stereoscopic displays are in fact consistent with the notion that myopes make less use of retinal defocus information (Cufflin, Mankowska, and Mallen 2007; Rosenfield and Abraham-Cohen 1999). However, the chain of causation between these deficits of myopic vision and the development of myopia still needs to be established. It might in fact be the case that inaccuracies of accommodation may cause long-term blur on the retina, which in turn could be responsible for the failure of emmetropization. If this were the case, stereoscopic displays which decouple the link between vergence and accommodation may play an adverse effect on the normal development of a healthy, emmetropic visual system.

Nevertheless, in Chapter 8, I have shown an example of how simulated disparity may be employed to tap into the binocular recalibration processes of the oculomotor system. By introducing non-detectable changes in the disparity content of visual targets between saccades, I have shown that it is possible to induce the oculomotor system to continuously recalibrate both rapid saccades and slow vergence eye movements independently of one another, and that corrections can even occur in opposite directions in each eye. These findings have implications in the treatment of amblyopia and strabis-

mus, deficits in which the two eye fail to properly work together. One further avenue of investigation into these visual deficits is to simulate them in visually normal observers, as shown in Chapter 9. By simulating monocular and binocular visual impairments, I have found that the neural linkage between the motor systems controlling the eyes and hands exists for eye and hand movements in three dimensions and is likely resilient to visual impairment. To then investigate the neural processing underlying the perception of motion, in Chapter 10 I have presented a neural model for estimating motion and self motion from optic flow. By directly comparing the behavioural performance of the model to that of human observers, I have shown how the retino-cortical lossy transformation of the primate visual system shapes the perception of motion in the peripheral visual field.

The work and findings I have presented in Chapters 3 through 10 raise an array of new research questions in basic, applied and clinical vision science, which I discuss in the next, and final, Chapter 15. The novel tools I have studied in Chapters 11 and 12 might be of great value in pursuing these new avenues of investigations. In Chapter 11 I have shown that the Tobii EyeX low-cost eye tracker is ideal for gathering preliminary eye movement data, pilot testing novel ideas, and could be easily integrated in commercial and clinical applications. In Chapter 12 I have demonstrated that the Oculus Rift head mounted display provides great levels of immersion with little evidence of simulator sickness symptoms. The Oculus Rift could thus be employed in clinical applications such as virtual reality exposure therapy, and in basic research for example investigating the role of self motion and action in visual perception.

# Chapter 14

## Limitations

The body of work presented in this thesis spans a wide range of topics concerning visual perception and oculomotor behaviour. Thus, I have necessarily employed a wide range of methodologies and experimental approaches. A series of caveats should be noted about the different statistical and experimental methods employed throughout this work.

In this thesis different test for assessing statistical significance have been employed, and the typical threshold value of  $p < 0.05$  has been employed throughout. Having set this threshold a priori, all results were discussed as significant if the statistical tests yielded  $p < 0.05$  and as not significant if the tests yielded  $p > 0.05$ . However, the 0.05 threshold is arbitrary; a p-value of 0.05 can be informally taken as meaning that, if a proposed model for the data is true (for example the null hypothesis that the means of two groups are the same) a result as extreme or more extreme than the one observed (i.e. incongruent with the model) will occur 1 in 20 times by chance. The American Statistical Association (ASA) released a statement in 2016 (Wasserstein

and Lazar 2016) reminding researchers of which principles should be applied when employing hypothesis testing and p-values, with particular focus on the notions that p-values do not measure the size of an effect. Thus, even though in this work I have employed the typical 0.05 threshold for statistical significance, I have also endeavored to present and discuss results and patterns of data in the context of the existing literature.

In Chapter 4 for example, I observed a statistically significant reduction in the time necessary to fuse stereoscopic images when peripheral blur was present at high uncrossed disparities. A similar pattern was also observed at high crossed disparities, however the statistical test did not reach the 0.05 threshold for significance. Whereas in Chapter 4 I strictly reported these results, it is likely that a benefit of peripheral blur was present at both crossed and uncrossed disparities, but showed up as statistically significant only at uncrossed disparities. As I discuss in Chapter 4, an asymmetry between fusion of crossed and uncrossed disparities is consistent with larger fusional ranges for crossed disparities than uncrossed disparities (Yeh and Silverstein 1990).

A separate caveat regarding the statistical analyses employed in Chapter 4 is that time to perceptual fusion data are always positive and skewed, as is typical of time to an event data. I employed ANOVA analysis on these data. However, a more appropriate statistical method for analyzing the time until an events occurs could have been survival analysis (Miller Jr 2011). In the experiment described in Chapter 4, I allowed participants unlimited time to report perceptual fusion. Had I restricted the trial time to a fixed amount, survival analysis would have also been able to deal with censored data, i.e.



instances in which the event of interest does not occur.

An additional observation concerns the number of human participants tested in each study. In psychophysics it is common to test small numbers of subjects, as few as 2 (Blakemore and Sutton 1969). Even though data from 2 subjects may appear to be too few, it is important to consider that for each subject hundreds if not thousands of data points are often collected. The typical psychophysical approach is in fact to obtain highly reliable data from extensive testing of few subjects. This is because psychophysics often investigates effects and mechanisms thought to be fundamental and common to all humans. To this point, it has been shown that as few as 5 subjects all showing an effect are sufficient to prove that the effect exists in a majority of the population (Anderson and Vingrys 2001). For these reasons the experiments presented in Chapters 3, 4, 8, 9 and 10 employed small numbers of subjects. In the other experimental chapters instead I employed larger numbers of subjects. In Chapters 5, 6 and 7, I required a greater number of subjects since I compared visual perception between different populations. In Chapters 11 and 12, I instead assessed the usability of commercial devices in scientific research with human subjects, and it was necessary to make certain the devices performed similarly across a larger sample of the population. Additional instances in which it might be useful to employ larger numbers of observers are psychophysical studies investigating individual differences in perception. For example, the recent internet phenomenon regarding the photograph named “the dress” highlighted that different portions of the population experienced very different colour percepts when viewing the same visual stimulus. Investigating these individual differences provides fundamental insights into how

the brain encodes and incorporates prior knowledge about the visual environment into the computations that lead to visual perception (Lafer-Sousa, Hermann, and Conway 2015; Gegenfurtner, Bloj, and Toscani 2015). An alternative approach to extensively testing a small number of subjects could also be that of testing a larger sample of the population whilst requesting few trials from each individual. These data could then be analyzed with Generalized Linear Mixed Models (GLMMs) which provide a framework in which it is possible to obtain reliable population-level performance metrics without discarding subject-specific variability (Moscatelli, Mezzetti, and Lacquaniti 2012).

A final consideration regarding the statistics employed in this thesis concerns the analyses of questionnaire data described in Chapter 12. Ordinal data such as the Likert scale data discussed in Chapter 12 are numerical data in which there is a specific order but the distance between two consecutive numbers is not known to be constant throughout the scale. Because the distances between consecutive values might not be equal, numerical operations on these data might not be valid and computing summary statistics such as the mean or the standard deviation might not be appropriate. In Chapter 12 I calculated a multiple linear regression to assess the relationship between Heart Rate Increment (dv) and Questionnaire Scores (predictors). Because I was interested in the simple question of whether or not a relationship between Heart Rate Increment and Questionnaire Scores existed, I calculated a multiple regression in spite of the fact that this required numerical operations on ordinal data. I deemed that the Likert data from the questionnaires in Chapter 12 could be reasonably treated as interval data given the nature

of the questions asked. However, a more appropriate and in-depth assessment of these data could be performed using Rasch analysis (Rasch 1960). A Rasch model could be employed to assess which questionnaire items could be combined into a total score which would determine the level of immersivity of each observer using VR equipment and whether this score is related to the level of physiological activation measured via heart rate monitoring.

In addition to the observations regarding statistical methodology reported above, readers should also be aware of certain observations regarding the experimental methodology employed in this thesis.

One experimental parameter which I discuss throughout the thesis but that I do not systematically control for is pupil size. Pupil size is an important parameter for the topics discussed in this thesis because it contributes to optical blur: small pupils produce little blur and extend the depth of field but allow less amounts of light into the eye, whereas larger pupils let more light into the eye but produce greater amounts of dioptric blur. Furthermore pupil size varies with accommodation: pupils widen when our eyes accommodate to near objects, whereas pupils constrict when our eyes accommodate to far distances. For this reason, in Chapter 4 I discussed that since pupil size appeared to vary with presented disparity it was likely that observers varied their accommodation throughout the experimental sessions and this would have induced additional blur to that specifically manipulated in the experiment. In the modelling presented Chapter 5, I imposed a small pupil size, which is plausible for outdoors light levels. I show that even when pupil size is the same, there are differences in peripheral blur distributions for outdoors and indoors scenes in the eyes of myopic and emmetropic ob-

servers. Incorporating into the modelling pupils that vary in size according to the light levels of outdoors and indoors natural scenes might modulate the magnitude of these differences. In Chapters 6 and 7, I compared perceptual and oculomotor performance between myopes and emmetropes and I did not control for pupil size. This was purposefully chosen because I wanted to assess how the different optics of myopic and emmetropic observers, of which pupil size is a component, affected visual perception and oculomotor control. Finally in Chapter 9, I discuss how pupil size affects the significance of my findings with regards to the general population. Thus, in future extensions of the work presented in this thesis (see next and final Chapter 15) pupil size will necessarily be an experimental parameter of great interest.

The pupil not only determines the amount of blur, but also determines the shape of the optical transfer function that describes how different spatial frequencies are handled by the optics of the eye. In this thesis I predominantly employ Gaussian filtering to simulate blur. However, the human pupil is a circular aperture that produces phase reversals in the retinal image, and the resulting blur can be better approximated using sinc filtering in the frequency domain (Murray and Bex 2010). If humans adapt to their own optics, it will be important to extend the findings presented in this thesis using different kinds of blur, other than Gaussian blur, that better approximate observer’s own optics. This could be accomplished using testing rigs capable of producing defocus blur through an observer’s own optics (Akeley et al. 2004; Love et al. 2009; Sebastian, Burge, and Geisler 2015), as I discuss in the next Chapter.

An additional optical parameter of interest for the investigations pre-

sented in this thesis is refractive correction with spectacles. The modelling presented in Chapter 5 assumes observers are best corrected with contact lenses and would have to be further complicated to account for the optical minification produced by spectacles (see Knapp's Law) (Chui et al. 2005; Colletta and Watson 2006). Additionally, the findings from Chapter 8 regarding unequal saccadic adaptation in our two eyes provide interesting predictions regarding the use of anisometric and multifocal spectacles. Specifically, my findings predict that observers should adapt their eye movements to their own spectacles, and that when wearing glasses with different prescription in the two eyes, observers' eye movements should adapt asymmetrically. This prediction is indeed confirmed by data from long-term anisometric spectacle lens wearers (Erkelens, Collewijn, and Steinman 1989; Lemij and Collewijn 1991a) and from observers subjected to short-term wearing of anisometric spectacles (Lemij and Collewijn 1991b).

A separate methodological caveat concerns the findings in Chapter 9 regarding pursuit and vergence tracking eye movements. Horizontal pursuit tracking accuracy was found to be significantly greater than vergence pursuit tracking accuracy. The reasons for this difference are manifold. First, vergence eye movements are likely more inaccurate than version eye movements because they require the additional computational step of estimating binocular disparity. However, the difference between vergence and pursuit tracking accuracy may also be driven, at least in part, by methodological limitations regarding vergence measurements. Firstly, measurement noise for each eye is independent and its magnitude is large relative to the small size of vergence eye movements. Secondly, vergence measures the distance between the eyes

and so measurement noise is additive, summing the noise from the two gaze measurements. Instead, version measures the average gaze positions of both eyes, which reduces the noise. Thus, vergence measurements will intrinsically be more noisy than horizontal pursuit measurements. Furthermore, the latency in the experimental setup is driven primarily by the screen refresh rate. Since I used a shutter-glasses stereoscopic 3D display, the stimulus in each eye was updated on alternate video frames. During pursuit tracking, latency can be estimated at 1 frame, approximately 8 ms, by averaging eye position. During vergence tracking, however, the latency will be 2 frames, i.e. 16 ms. Thus the observed difference between vergence and pursuit tracking accuracy may be driven in part by methodological limitations of the experimental setup. Nevertheless, the main pattern of results and conclusions concerning eye-hand coupling in conditions of visual uncertainty and in 3D are unlikely to be affected by these limitations.

In Chapter 10 I had human observers estimate the focus of radial motion from expanding or contracting dead leaves stimuli. The stimuli simulated the pattern of optic flow observers would experience when approaching or receding from a flat surface, similarly to what was performed by Bex and Falkenberg (2006). In the discussion, I commented on the fact that the data are noisy and individual estimates of Internal Noise and Sampling Efficiency are highly variable in both studies. This could in part be due to the fact that heading judgments using approach to a surface stimuli are a special case that constitutes a notoriously difficult task for human observers (Warren and Hannon 1990). Further testing with stimuli simulating an approach to a cloud of dead leaf elements at different distances from the observer might

provide cleaner data. These stimuli would also be well suited to test future iterations of the model described in Chapter 10 which could include disparity processing for the perception of motion in depth, which I further discuss in the next Chapter.

Finally, throughout this thesis I have attempted to study the visual system using naturalistic stimuli. In Chapters 3 and 4, I employed real photographs of natural scenes, in Chapter 5, I employed range images of real outdoors and indoors scenes, and in most other chapters I employed dead leaves stimuli which have a similar  $1/f$  spatial frequency spectrum and contrast range of natural images, and are textured with occlusions and edges at a variety of orientations (Bordenave, Gousseau, and Roueff 2006; Lee, Mumford, and Huang 2001). Natural and naturalistic stimuli have advantages and disadvantages. Natural stimuli are not as easily controllable as simplified artificial stimuli. In artificial stimuli the experimenter can parametrically vary low-level image properties such as contrast, spatial frequency and orientation and then test the limits and performance of visual perception as each parameter varies individually. Conversely, in complex natural scenes all image properties continuously vary and co-vary, and the experimenter has to assess which image features more closely correlate with visual perception. Because the experimenter is not directly controlling the stimulus, it is easier to fail to control for unexpected confounds in natural stimuli. However, it can be argued that testing with impoverished laboratory stimuli will not provide the full explanation of how the visual system works, because the visual system has developed and has been shaped by evolution to work within the complexity of the natural environment. Indeed it has been suggested that

the perception of visual space is determined by priors based on the probability distribution of real-world sources of retinal images (Howe and Purves 2002). Thus a comprehensive approach could be that of approaching each question about the visual system from both perspectives. Experimenters should verify that conclusions based on simple artificial stimuli generalize to natural viewing, and that correlational findings with natural images can be replicated with highly controllable laboratory stimuli.



## Chapter 15

# Suggestions for Future Research

Visual perception and oculomotor behaviour are tightly interlinked in an elegant yet highly complex system that has evolved to allow us to explore and interact with our environment. Our central vision inspects fine details in our visual environment, while our low resolution and blurry peripheral vision selects targets for future inspection. To shift our high resolution foveae onto newly selected targets at different locations and distances, our oculomotor systems plans and executes complex sequences of saccadic, vergence and pursuit eye movements. To keep sharply focused images onto our foveae, our accommodation system constantly changes the focal power of our crystalline lenses. As our visual system is developing, the shape of our eyes determines our perceptual experience of the visual world, and the visual signals throughout the retina determine the rate of growth and the shape of our eyes.

To fully understand the complex interplay between all the factors at work during visual development as described in this thesis, novel display technology might be necessary. A volumetric display capable of correctly and independently driving vergence, accommodation, and pupil diameter could be employed to tease different factors apart (Akeley et al. 2004; Watt, Akeley, Girshick, et al. 2005; Love et al. 2009; Liu and Hua 2010; MacKenzie, Hoffman, and Watt 2010; Sebastian, Burge, and Geisler 2015). First, healthy oculomotor behaviour in real world should be compared to oculomotor behaviour in a volumetric display that provides correct depth and blur cues, and allows for natural and correctly linked eye movements and accommodation in 3D. Then, when display technology is capable of driving healthy oculomotor behaviour, performance which this display should be compared to a progressively impoverished display setup. Individually removing accommodation, blur distributions and stereoscopic disparity would allow us to understand the relative contributions of each factor to healthy oculomotor behaviour, and which factors are most likely to negatively impact normal oculomotor development.

Since in individuals with refractive errors blur perception, accommodation, and vergence eye movements are hindered (see Part 2, Chapters 5, 6, and 7), the causal relationship between blur perception, correct accommodation, and correct vergence eye movements should be established to understand which factor might be more directly responsible for incorrect emmetropization. Blur perception should be evaluated with stabilized and destabilized accommodation. Accommodation should be evaluated with correct and incorrect focus and depth cues. Vergence eye movements should be assessed

with correct and incorrect focus cues and accommodation. All of these subsystems should be investigated in prospective/longitudinal studies in children at risk of developing myopia. This would allow researches to pinpoint which deficits come first and cause the others, or if they all develop concurrently. Only then will it be possible to develop viable treatment and preventative strategies that target the first causal step in the chain of deficits that lead to incorrect eye growth and the development of refractive errors.

The chain of events that leads to refractive errors could be initiated by deficits in the accommodation system for example. The distributions of dioptric depths in indoors environments, which are more varied than the distributions of dioptric depths outdoors (Flitcroft 2012), could place a strain on the accommodative system of a child, and lead to inaccuracies of accommodation which would in turn lead to blur on the retina and prevent normal emmetropization (Goss 1991; Gwiazda, Thorn, and Held 2005; Gwiazda et al. 2004; Hampson, Cufflin, and Mallen 2013; McBrient and Millodot 1986; Millodot 2015). Thus viable treatment strategies could include encouraging children to spend more time outdoors (Jones et al. 2007; Rose et al. 2008; Dirani et al. 2009). However, since children spend increasing amounts of time looking at monitors (Rideout, Foehr, and Roberts 2011), an alternative treatment strategy could be to create display technology which provides the possibility of accommodating at depths other than that of the monitor (Akeley et al. 2004; Watt, Akeley, Girshick, et al. 2005; Love et al. 2009; Liu and Hua 2010; MacKenzie, Hoffman, and Watt 2010; Sebastian, Burge, and Geisler 2015), and provide a range of dioptric depths that matches that of outdoors environments (Maiello, Bex, and Vera-Diaz 2014).

A different chain of events that could explain the development of refractive errors could be that the mechanisms driving eye growth may be attempting to obtain a retinal shape that produces blur distributions that match those experienced by a healthy eye in outdoors viewing. Since indoors and outdoors blur distributions are markedly different (Chapter 5), children spending excessive time indoors could lead to incorrect, myopic eye growth, which in turn would lead to excessive retinal blur, which could impact negatively on blur sensitivity and in turn destabilize accommodation and vergence. Thus a viable treatment strategy could be to provide children with display technology that mimics outdoors distributions of blur (see Chapter 4, and Maiello et al. 2014; Maiello, Bex, and Vera-Diaz 2014).

An important step in devising such technology would be to assess whether rendered blur produces the same effects as blur due to an observer's own optics. It is known that the perception of rendered blur varies with the digital filter employed to create blur, and that sinc blur is likely the best approximation of optical blur since it introduces periodic phase reversals similar to those due to the modulation transfer function of an optical system with a circular aperture such as the human pupil (Murray and Bex 2010). Studies in animal models such as chicks could be conducted to verify whether blur rendered on displays using different filtering techniques has the same effects on eye growth as blur due to lenses (Wallman and Winawer 2004).

In Part 2 of this thesis I discussed how eye shape might affect visual perception, and I have shown that myopes, who have elongated retinae, have worse peripheral contrast sensitivity than emmetropes. However, in this work I did not directly assess the eye shape of the observers who participated in

the study. To establish a clear link between retinal shape and perception, eye shape should be measured directly, for example using OCT imaging (Atchison et al. 2005). This step would be of particular importance when investigating other potential perceptual consequences of elongated eyes. For example, visual function in the periphery is primarily limited not by acuity or contrast sensitivity but by visual crowding, the phenomenon by which visible targets become harder to identify in clutter. Because of the stretched myopic retina, objects projected onto the peripheral retinal surface of myopic eyes subtend the same retinal angle, but are spaced farther apart than on an emmetropic retina. In preliminary data (Carroll et al. 2016) I have found that spatial interference zones in myopic eyes are smaller than in emmetropic eyes. These data would suggest that the elongated retinal surface of myopic eyes may provide a functional benefit by extending the spacing within which peripheral targets are crowded. Thus it is possible that multiple aspects of visual perception are tied to eye shape. Once this link is established, it might even be possible to estimate an individual's eye shape from his/her performance at perceptual tasks in the visual periphery.

In Chapter 9 I have shown that oculomotor performance and hand-eye coordination deteriorate with 1-3 diopters of simulated blur. This finding warrants the investigation of hand-eye coordination and oculomotor control in cross-sectional studies spanning the range of refractive errors present in the general population, in order to identify the critical refraction at which oculomotor control begins to break down in observers who have adapted to their optical deficits. As mentioned previously, longitudinal studies in children could instead help identify the causal relationships between refractive

errors and visual and oculomotor deficits.

In binocular vision deficits such as strabismus, convergence insufficiency and amblyopia the eyes fail to work together. Treating strabismus with surgery often does not work, the deviated eye tends to return to its deviated position after surgery (Fresina and Campos 2014). Furthermore, restoring visual function in the amblyopic eye of a strabismic amblyope with patching therapy, which does not promote binocular cooperation, might result in diplopia when the two eyes are not aligned. However the visual system has the capability for plasticity and recalibration, which we may be able to tap into when developing strategies to treat and prevent these disorders of binocular vision. Thus, an innovative treatment strategy might be to gradually teach the eyes to work together, in terms of both perceptual fusion and oculomotor alignment. Novel dichoptic treatment strategies (Hess, Mansouri, and Thompson 2010; To et al. 2011; Knox et al. 2012; J. Li et al. 2013; Tailor et al. 2015; Tailor et al. 2016) might thus be coupled with binocular oculomotor recalibration, using the tools and findings described in Chapter 8. I have also shown that there exists a link between the hand and eye motor systems, and that this link persists even with monocular and binocular simulated impairments. Thus it may be possible to exploit this linkage to enhance binocular rehabilitation treatments. Tracking self generated hand motion could stabilize oculomotor performance, which could in turn facilitate binocular perceptual cooperation.

Within this context, a lacuna in the literature on oculomotor control in three dimensional space concerns the behaviour of vergence eye movements with unbalanced visual input. Thus, in preliminary data (Maiello, Gibaldi,

et al. 2016) I have been finding that interocular differences in contrast impair both vergence and depth perception, whereas interocular differences in luminance do not impair vergence but strongly impair depth perception. These findings can be described with a computational neural model of vergence control (Gibaldi et al. 2012), based on a network of V1 simple and complex cells. Similarly, in pilot data (Maiello, Chessa, Bex, et al. 2016) I have been showing that an algorithm for disparity that mimics the processing stages occurring in primary visual cortex performs comparably to human observers at depth discrimination tasks. These neural models are based on similar architectures as the one described in Chapter 10 for processing of optic flow information. Thus, these different modules could be combined to develop a functional neural model of perception of depth, motion and oculomotor control in three dimensional space. Such neural models would be valuable for understanding how both the healthy and impaired visual systems operate. By selectively lesioning the neural model at different processing stages, we may attempt to recreate binocular visual impairments. This would help researchers hypothesize at what processing stage deficits such as amblyopia first develop, as well as devise treatment strategies targeted for specific neural pathways.

Novel low-cost technologies, such as those investigated in Chapters 11 and 12 are likely to play an important role in future research on the binocular visual system. As these technologies develop and new ones are released, it will be important to continuously evaluate their performance with respect to basic, clinical, and applied research. In the toolkit I have contributed to developing when assessing the Tobii EyeX eye tracker for example, I have in-

cluded methods and software tools to evaluate future devices as well (Gibaldi et al. 2016). Tobii for instance is in the process of releasing a new low-cost eye tracker, the Tobii Eye Tracker 4C. Head mounted displays such as the HTC Vive, the Samsung Gear VR, the OSVR Open Source Head-mounted Display will also need to be assessed in detail. Some of these novel low cost solutions such as the Qualcomm Snapdragon VR820 and the FOVE Eye Tracking VR Headset even integrate head mounted display technology with eye tracking. Hand motion tracking may be included in some of these devices using the low cost Leap Motion hand tracker employed in Chapter 9.

These novel immersive technologies could open the doors to new ways of investigating the mind. Human visual experience is seemingly rich and detailed, despite the fact that only the very center of our retina, the high-resolution fovea, is able to analyze fine structure. As mentioned previously, since visual resolution decreases with distance from the fovea, observers must first select targets with low-resolution peripheral vision, then make coordinated saccade and, if the object is at a different depth, vergence binocular eye movements that bring the fovea of each eye onto that target. During each fixation, a gradient of visual information falling away from the fovea is available to the observer, however not all of this information reaches awareness. For example, humans are capable of discriminating extremely fine depth differences based on stereoscopic disparity or optical blur differences between pairs of experimental images. However, we are only able to perceive 3 or 4 different depth simultaneously based on stereoscopic (Tsirlin, Allison, and Wilcox 2008) or blur (Taylor and Bex 2015) information. Thus, only a fraction of the resolvable information is processed and stored by the visual system each



fixation. To maintain a rich and detailed neural representation of the visual world, a lossy internal model of the world must be built up and integrated across multiple fixations. For each object in the world, the observer's internal model of the world will encode a range of object attributes, including identity (what is the object?), 3D spatial location (where is the object?), and object motion (how is the object moving?). The structure of this internal representation will be shaped further by the observer's actions within the world and their interactions with the objects within the world. Novel virtual reality technology could thus allow us to investigate the construction, structure and the capacity of this internal model with highly controllable virtual stimuli yet in a naturalistic setting where subjects may act and interact with the visual environment. These new levels of understanding of how humans visually perceive the world could not only further our knowledge of the brain, but also shape the evolution of VR technology. By providing a mechanistic description of visual scene perception, we could begin developing VR rendering algorithms based not on models of the physical world, but on sophisticated models of visual perception and visual scene encoding.

# Bibliography

- Abbott, M. L., K. L. Schmid, and N. C. Strang. 1998. "Differences in the accommodation stimulus response curves of adult myopes and emmetropes." *Ophthalmic and Physiological Optics* 18 (1): 13–20.
- Adamovich, S. V., G. G. Fluet, E. Tunik, and A. S. Merians. 2009. "Sensorimotor training in virtual reality: a review." *NeuroRehabilitation* 25 (1): 29–44.
- Adelson, E. H., and J. R. Bergen. 1985. "Spatiotemporal energy models for the perception of motion." *Journal of the Optical Society of America A* 2:284–299.
- Adelson, E. H., and J. A. Movshon. 1982. "Phenomenal coherence of moving visual patterns." *Nature* 300 (5892): 523–525.
- Akeley, K., S. J. Watt, A. R. Girshick, and M. S. Banks. 2004. "A Stereo Display Prototype with Multiple Focal Distances." *ACM Transactions on Graphics* 23 (3): 804–813.
- Albano, J. E., and J. A. Marrero. 1995. "Binocular interactions in rapid saccadic adaptation." *Vision Research* 35 (23-24): 3439–3450.

- Allen, P. M., H. Radhakrishnan, S. Rae, R. I. Calver, B. P. Theagarayan, P. Nelson, E. Osuoben, A. Sailoganathan, H. Price, and D. J. O’Leary. 2009. “Aberration control and vision training as an effective means of improving accommodation in individuals with myopia.” *Investigative Ophthalmology & Visual Science* 50 (11): 5120–5129.
- Allison, R., I. Howard, and X. Fang. 2004. “The stimulus integration area for horizontal vergence.” *Experimental Brain Research* 156 (3): 305–313.
- Alvarez, T., J. Semmlow, W. Yuan, and P. Munoz. 2002. “Comparison of disparity vergence system responses to predictable and non-predictable stimulations.” *Current Psychology of Cognition* 21 (2/3): 243–262.
- Anderson, A. J., and A. J. Vingrys. 2001. “Small samples: does size matter?” *Investigative Ophthalmology & Visual Science* 42 (7): 1411–1413.
- Anderson, H. A., A. Glasser, R. E. Manny, and K. K. Stuebing. 2010. “Age-related changes in accommodative dynamics from preschool to adulthood.” *Investigative Ophthalmology & Visual Science* 51 (1): 614–622.
- Anderson, T. J., I. Jenkins, D. J. Brooks, M. Hawken, R. S. Frackowiak, and C. Kennard. 1994. “Cortical control of saccades and fixation in man A PET study.” *Brain* 117 (5): 1073–1084.
- Applegate, R., G. Hilmantel, and H. Howland. 1997. “Area under log contrast sensitivity function: a concise method of following changes in visual performance.” *OSA Technical Digest Series* 1:98–101.

- Artal, P., L. Chen, E. J. Fernández, B. Singer, S. Manzanera, and D. R. Williams. 2004. “Neural compensation for the eye’s optical aberrations.” *Journal of Vision* 4 (4): 4–4.
- Ashby, R., A. Ohlendorf, and F. Schaeffel. 2009. “The effect of ambient illuminance on the development of deprivation myopia in chicks.” *Investigative Ophthalmology & Visual Science* 50 (11): 5348–5354.
- Aslin, R. N. 1977. “Development of binocular fixation in human infants.” *Journal of Experimental Child Psychology* 23 (1): 133–150.
- Atchison, D. A., N. Pritchard, K. L. Schmid, D. H. Scott, C. E. Jones, and J. M. Pope. 2005. “Shape of the retinal surface in emmetropia and myopia.” *Investigative Ophthalmology & Visual Science* 46 (8): 2698–2707.
- Atchison, D. A., K. L. Schmid, and N. Pritchard. 2006. “Neural and optical limits to visual performance in myopia.” *Vision Research* 46 (21): 3707–3722.
- Atkinson, J., O. Braddick, and K. Moar. 1977. “Development of contrast sensitivity over the first 3 months of life in the human infant.” *Vision Research* 17 (9): 1037–1044.
- Ayeni, A., W. Harrison, and P. J. Bex. 2015. “Filling-in of Kanizsa-style illusory figures is under top-down control.” *Journal of Vision* 15 (12): 532–532.
- Badcock, D. R., and C. M. Schor. 1985. “Depth-increment detection function for individual spatial channels.” *Journal of the Optical Society of America A* 2 (7): 1211–1215.

- Bahill, A., A. Brockenbrough, and B. Troost. 1981. "Variability and development of a normative data base for saccadic eye movements." *Investigative Ophthalmology & Visual Science* 21 (1): 116–25.
- Bailenson, J. N., N. Yee, S. Brave, D. Merget, and D. Koslow. 2007. "Virtual interpersonal touch: expressing and recognizing emotions through haptic devices." *Human–Computer Interaction* 22 (3): 325–353.
- Balas, B., L. Nakano, and R. Rosenholtz. 2009. "A summary-statistic representation in peripheral vision explains visual crowding." *Journal of Vision* 9 (12): 13.
- Baloh, R., A. Sills, W. Kumley, and V. Honrubia. 1975. "Quantitative measurement of saccade amplitude, duration, and velocity." *Neurology* 25 (11): 1065–1065.
- Bando, T., A. Iijima, and S. Yano. 2012. "Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review." *Displays* 33 (2): 76–83.
- Banks, M. S. 1980. "The development of visual accommodation during early infancy." *Child Development*: 646–666.
- Banks, M., and P. Salapatek. 1978. "Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants." *Investigative Ophthalmology & Visual Science* 17 (4): 361–365.
- Banton, T., and D. M. Levi. 1991. "Binocular summation in vernier acuity." *Journal of the Optical Society of America A* 8 (4): 673–680.

- Barlow, H. B. 1956. “Retinal noise and absolute threshold.” *Journal of the Optical Society of America* 46 (8): 634–639.
- Beardsley, S. A., and L. M. Vaina. 2001. “A laterally interconnected neural architecture in MST accounts for psychophysical discrimination of complex motion patterns.” *Journal of Computational Neuroscience* 10 (3): 255–280.
- Beers, A., and G. Van Der Heijde. 1994. “In vivo determination of the biomechanical properties of the component elements of the accommodation mechanism.” *Vision Research* 34 (21): 2897–2905.
- Beers, A., and G. Van der Heijde. 1996. “Age-related changes in the accommodation mechanism.” *Optometry & Vision Science* 73 (4): 235–242.
- Behrens, F., M. MacKeben, and W. Schröder-Preikschat. 2010. “An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters.” *Behavior Research Methods* 42 (3): 701–708.
- Bellucci, R. 2005. “Multifocal intraocular lenses.” *Current Opinion in Ophthalmology* 16 (1): 33–37.
- Benjamini, Y., and Y. Hochberg. 1995. “Controlling the false discovery rate: a practical and powerful approach to multiple testing.” *Journal of the Royal Statistical Society. Series B*: 289–300.
- Bernhaupt, R., K. Isbister, and S. De Freitas. 2015. “Introduction to this Special Issue on HCI and Games.” *Human–Computer Interaction* 30 (3-4): 195–201.

- Berntsen, D. A., L. T. Sinnott, D. O. Mutti, K. Zadnik, C. S. Group, et al. 2011. "Accommodative lag and juvenile-onset myopia progression in children wearing refractive correction." *Vision Research* 51 (9): 1039–1046.
- Berton, F., G. Sandini, and G. Metta. 2006. "Encyclopedia of Sensors." Chap. Anthropomorphic visual sensors, 1–16. American Scientific Publishers.
- Bex, P. J., and W. Makous. 2002. "Spatial frequency, phase, and the contrast of natural images." *Journal of the Optical Society of America A* 19 (6): 1096–1106.
- Bex, P. J., A. Ayeni, and E. Wiecek. 2015. "Word and Sentence Level Spatial Information In Reading." *Journal of Vision* 15 (12): 911–911.
- Bex, P. J., and H. K. Falkenberg. 2006. "Resolution of complex motion detectors in the central and peripheral visual field." *Journal of the Optical Society of America A* 23 (7): 1598–1607.
- Birch, E. E. 2013. "Amblyopia and binocular vision." *Progress in Retinal and Eye Research* 33:67–84.
- Blakemore, C., and P. Sutton. 1969. "Size Adaptation: A New Aftereffect." *Science* 166 (3902): 245–247.
- Bolduc, M., and M. D. Levine. 1998. "A review of biologically motivated space-variant data reduction models for robotic vision." *Computer Vision and Image Understanding* 69 (2): 170–184.

- Bonmassar, G., and E. L. Schwartz. 1997. "Space-variant fourier analysis: the exponential chirp transform." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19 (10): 1080–1089.
- Bonnen, K., J. Burge, J. Yates, J. Pillow, and L. K. Cormack. 2015. "Continuous psychophysics: Target-tracking to measure visual sensitivity." *Journal of Vision* 15 (3): 14.
- Bordenave, C., Y. Gousseau, and F. Roueff. 2006. "The dead leaves model: a general tessellation modeling occlusion." *Advances in Applied Probability*: 31–46.
- Botella, C., S. Quero, R. Baños, C. Perpiñá, P. A. García, and G. Riva. 2004. "Virtual reality and psychotherapy." *Studies in Health Technology and Informatics* 99:37.
- Bouchard, S. 2011. "Could virtual reality be effective in treating children with phobias?" *Expert Review of Neurotherapeutics* 11 (2): 207–213.
- Bradley, N. A., and M. D. Dunlop. 2005. "Toward a multidisciplinary model of context to support context-aware computing." *Human-Computer Interaction* 20 (4): 403–446.
- Brainard, D. H. 1997. "The Psychophysics Toolbox." *Spatial Vision* 10:433–436.
- Bridgeman, B., D. Hendry, and L. Stark. 1975. "Failure to detect displacement of the visual world during saccadic eye movements." *Vision Research* 15 (6): 719–722.



- Brodie, S. E. 1987. “Photographic calibration of the Hirschberg test.” *Investigative Ophthalmology & Visual Science* 28 (4): 736–742.
- Browning, N. A., S. Grossberg, and E. Mingolla. 2009. “A neural model of how the brain computes heading from optic flow in realistic scenes.” *Cognitive Psychology* 59 (4): 320–356.
- Bullimore, M. A., B. Gilmartin, and J. M. Royston. 1992. “Steady-state accommodation and ocular biometry in late-onset myopia.” *Documenta Ophthalmologica* 80 (2): 143–155.
- Burge, J., and W. S. Geisler. 2011. “Optimal defocus estimation in individual natural images.” *Proceedings of the National Academy of Sciences* 108 (40): 16849–16854.
- Burt, P., and E. Adelson. 1983. “The Laplacian Pyramid as a Compact Image Code.” *IEEE Transactions on Communications* 31:532–540.
- Campbell, F. W. 1957. “The depth of field of the human eye.” *Journal of Modern Optics* 4 (4): 157–164.
- Carandini, M. 2012. “Area V1.” *Scholarpedia* 7 (7): 12105.
- Carney, L. G., J. C. Mainstone, and B. A. Henderson. 1997. “Corneal topography and myopia. A cross-sectional study.” *Investigative Ophthalmology & Visual Science* 38 (2): 311.
- Carroll, S., G. Maiello, W. Harrison, and P. J. Bex. 2016. “Myopic eyes see better in a crowd.” *Journal of Vision* 16 (12): 233–233.

- Charman, N. W. 2011a. “Myopia: its prevalence, origins and control.” *Ophthalmic and Physiological Optics* 31 (1): 3–6.
- . 2011b. “Myopia, posture and the visual environment.” *Ophthalmic and Physiological Optics* 31 (5): 494–501.
- Charman, N. W., and R. Hema. 2010. “Peripheral refraction and the development of refractive error: a review.” *Ophthalmic and Physiological Optics* 30 (4): 321–338.
- Chaturvedi, V., and J. A. Van Gisbergen. 1997. “Specificity of saccadic adaptation in three-dimensional space.” *Vision Research* 37 (10): 1367–1382.
- Chen, J., M. Valsecchi, and K. R. Gegenfurtner. 2016. “LRP predicts smooth pursuit eye movement onset during the ocular tracking of self-generated movements.” *Journal of Neurophysiology*.
- Chessa, M., F. Solari, and S. P. Sabatini. 2013. “Adjustable linear models for optic flow based obstacle avoidance.” *Computer Vision and Image Understanding* 117 (6): 603–619.
- Chessa, M., G. Maiello, P. J. Bex, and F. Solari. 2016. “A space-variant model for motion interpretation across the visual field.” *Journal of Vision* 16 (2): 12–12.
- Chessa, M., G. Maiello, A. Borsari, and P. J. Bex. 2016. “The Perceptual Quality of the Oculus Rift for Immersive Virtual Reality.” *Human-Computer Interaction*, no. in press.

- Chessa, M., S. Sabatini, F. Solari, and F. Tatti. 2011. "A Quantitative Comparison of Speed and Reliability for Log-Polar Mapping Techniques." In *Computer Vision Systems*, edited by J. Crowley, B. Draper, and M. Thonnat, 6962:41–50. Lecture Notes in Computer Science.
- Choi, M., S. Weiss, F. Schaeffel, A. Seidemann, H. C. Howland, B. Wilhelm, and H. Wilhelm. 2000. "Laboratory, clinical, and kindergarten test of a new eccentric infrared photorefractor (PowerRefractor)." *Optometry & Vision Science* 77 (10): 537–548.
- Christman, S. 1990. "Effects of luminance and blur on hemispheric asymmetries in temporal integration." *Neuropsychologia* 28 (4): 361–374.
- Chui, T. Y., M. K. Yap, H. H. Chan, and L. N. Thibos. 2005. "Retinal stretching limits peripheral visual acuity in myopia." *Vision Research* 45 (5): 593–605.
- Ciuffreda, K. J. 1991. "Accommodation and its anomalies." *Vision and visual dysfunction* 1:231–279.
- . 1998. "Accommodation, the pupil, and presbyopia." *Borish's Clinical Refraction*, ed. by W. Benjamin: 77–120.
- Ciuffreda, K. J., B. Wang, and B. Vasudevan. 2007. "Conceptual model of human blur perception." *Vision Research* 47 (9): 1245–1252.
- Clark, R. A., and J. L. Demer. 2002. "Effect of aging on human rectus extraocular muscle paths demonstrated by magnetic resonance imaging." *American Journal of Ophthalmology* 134 (6): 872–878.

- Cohen, M., and L. Ross. 1977. "Saccade latency in children and adults: Effects of warning interval and target eccentricity." *Journal of Experimental Child Psychology* 23 (3): 539–549.
- Coletta, N. J., and T. Watson. 2006. "Effect of myopia on visual acuity measured with laser interference fringes." *Vision Research* 46 (5): 636–651.
- Collewijn, H., C. J. Erkelens, and R. M. Steinman. 1995. "Voluntary binocular gaze-shifts in the plane of regard: dynamics of version and vergence." *Vision Research* 35 (23): 3335–3358.
- . 1997. "Trajectories of the human binocular fixation point during conjugate and non-conjugate gaze-shifts." *Vision Research* 37 (8): 1049–1069.
- Collins, T., and J. Wallman. 2012. "The relative importance of retinal error and prediction in saccadic adaptation." *Journal of Neurophysiology* 107 (12): 3342–3348.
- Corbetta, M. 1998. "Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems?" *Proceedings of the National Academy of Sciences* 95 (3): 831–838.
- Cornelissen, F. W., E. M. Peters, and J. Palmer. 2002. "The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox." *Behavior Research Methods, Instruments, & Computers* 34 (4): 613–617.

- Coubard, O. A. 2013. “Saccade and vergence eye movements: a review of motor and premotor commands.” *European Journal of Neuroscience* 38 (10): 3384–3397.
- Cruz-Neira, C., D. J. Sandin, and T. A. DeFanti. 1993. “Surround-screen projection-based virtual reality: the design and implementation of the CAVE.” In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, 135–142. ACM.
- Cufflin, M. P., A. Mankowska, and E. A. H. Mallen. 2007. “Effect of Blur Adaptation on Blur Sensitivity and Discrimination in Emmetropes and Myopes.” *Investigative Ophthalmology & Visual Science* 48 (6): 2932.
- Cummings, J. J., and J. N. Bailenson. 2016. “How immersive is enough? A meta-analysis of the effect of immersive technology on user presence.” *Media Psychology* 19 (2): 272–309.
- Cyr, G., and D. Fender. 1969. “The interplay of drifts and flicks in binocular fixation.” *Vision Research* 9 (2): 245–265.
- Dalmajer, E. 2014. *Is the low-cost EyeTribe eye tracker any good for research?* Report 2167-9843. PeerJ PrePrints.
- Davis, S., K. Nesbitt, and E. Nalivaiko. 2014. “A systematic review of cyber-sickness.” In *Proceedings of the 2014 Conference on Interactive Entertainment*, 1–9. ACM.

- Day, M., N. C. Strang, D. Seidel, L. S. Gray, and E. A. Mallen. 2006. "Refractive group differences in accommodation microfluctuations with changing accommodation stimulus." *Ophthalmic and Physiological Optics* 26 (1): 88–96.
- Deng, L., J. Gwiazda, and F. Thorn. 2010. "Children's refractions and visual activities in the school year and summer." *Optometry and Vision Science* 87 (6): 406.
- Deubel, H., B. Bridgeman, and W. X. Schneider. 1998. "Immediate post-saccadic information mediates space constancy." *Vision Research* 38 (20): 3147–3159.
- Deubel, H., W. X. Schneider, and B. Bridgeman. 1996. "Postsaccadic target blanking prevents saccadic suppression of image displacement." *Vision Research* 36 (7): 985–996.
- . 2002. "Transsaccadic memory of position and form." *Progress in Brain Research* 140:165–180.
- Deubel, H., W. Wolf, and G. Hauske. 1985. "Adaptive gain control of saccadic eye movements." *Human Neurobiology* 5 (4): 245–253.
- Dirani, M., Y.-H. Chan, G. Gazzard, D. M. Hornbeak, S.-W. Leo, P. Selvaraj, B. Zhou, T. L. Young, P. Mitchell, R. Varma, et al. 2010. "Prevalence of refractive error in Singaporean Chinese children: the strabismus, amblyopia, and refractive error in young Singaporean Children (STARS) study." *Investigative Ophthalmology & Visual Science* 51 (3): 1348–1355.

- Dirani, M., L. Tong, G. Gazzard, X. Zhang, A. Chia, T. L. Young, K. A. Rose, P. Mitchell, and S.-M. Saw. 2009. "Outdoor activity and myopia in Singapore teenage children." *British Journal of Ophthalmology* 93 (8): 997–1000.
- Dodge, R. 1903. "Five types of eye movement in the horizontal meridian plane of the field of regard." *American Journal of Physiology* 8 (4): 307–329.
- Dodgson, N. A. 2004. "Variation and extrema of human interpupillary distance." In *Stereoscopic Displays and Virtual Reality Systems XI*, 5291:36–46.
- Dorr, M., and P. J. Bex. 2013. "Peri-saccadic natural vision." *The Journal of Neuroscience* 33 (3): 1211–1217.
- Dowling, J. 2007. "Retina." *Scholarpedia* 2 (12): 3487.
- Duchowski, A. T., D. H. House, J. Gestring, R. I. Wang, K. Krejtz, I. Krejtz, R. Mantiuk, and B. Bazyluk. 2014. "Reducing visual discomfort of 3d stereoscopic displays with gaze-contingent depth-of-field." In *Proceedings of the ACM Symposium on Applied Perception*, 39–46. ACM.
- Duffy, C. J., and R. H. Wurtz. 1991. "Sensitivity of MST Neurons to Optic Flow Stimuli. I. A Continuum of Response Selectivity to Large-Field Stimuli." *Journal of Neurophysiology* 65:1329–1345.
- Duffy, C. J., and R. H. Wurtz. 1995. "Response of monkey MST neurons to optic flow stimuli with shifted centers of motion." *The Journal of Neuroscience* 15 (7): 5192–5208.

- Efron, B., and R. Tibshirani. 1993. *An introduction to the bootstrap*. New York: Chapman / Hall.
- Emoto, M., T. Niida, and F. Okano. 2005. "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television." *Journal of Display Technology* 1 (2): 328–340.
- Enright, J. T. 1992. "The remarkable saccades of asymmetrical vergence." *Vision Research* 32 (12): 2261–2276.
- Erkelens, C. J., H. Collewijn, and R. Steinman. 1989. "Asymmetrical adaptation of human saccades to anisometropic spectacles." *Investigative Ophthalmology & Visual Science* 30 (6): 1132–1145.
- Fan, D., E. Cheung, R. Lai, A. Kwok, and D. Lam. 2004. "Myopia progression among preschool Chinese children in Hong Kong." *Annals Academy of Medicine Singapore* 33 (1): 39–43.
- Felton, T. B., W. Richards, and R. A. Smith. 1972. "Disparity processing of spatial frequencies in man." *The Journal of Physiology* 225 (2): 349–362.
- Ferrer-Blasco, T., and D. Madrid-Costa. 2010. "Stereoacuity with simultaneous vision multifocal contact lenses." *Optometry & Vision Science* 87 (9): E663–E668.
- Fielder, A. R., and M. J. Moseley. 1996. "Does stereopsis matter in humans?" *Eye* 10 (2): 233–238.
- Fincham, E., and J. Walton. 1957. "The reciprocal actions of accommodation and convergence." *The Journal of Physiology* 137 (3): 488.



- Fioravanti, F., P. Inchingolo, S. Pensiero, and M. Spanio. 1995. "Saccadic eye movement conjugation in children." *Vision Research* 35 (23): 3217–3228.
- Fisher, R. A. 1915. "Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population." *Biometrika* 10 (4): 507–521.
- Fisher, S. K., and K. J. Ciuffreda. 1988. "Accommodation and apparent distance." *Perception* 17 (5): 609–621.
- Fleet, D. J., H. Wagner, and D. J. Heeger. 1996. "Neural encoding of binocular disparity: energy models, position shifts and phase shifts." *Vision Research* 36 (12): 1839–1857.
- Flitcroft, D. 1998. "A model of the contribution of oculomotor and optical factors to emmetropization and myopia." *Vision Research* 38 (19): 2869–2879.
- . 2012. "The complex interactions of retinal, optical and environmental factors in myopia aetiology." *Progress in Retinal and Eye Research* 31 (6): 622–660.
- Foulkes, A. J., S. Rushton, and P. A. Warren. 2013a. "Flow parsing and heading perception show similar dependence on quality and quantity of optic flow." *Frontiers in Behavioral Neuroscience* 7.
- . 2013b. "Heading recovery from optic flow: comparing performance of humans and computational models." *Frontiers in Behavioral Neuroscience* 7:53.

- Fox, P. T., J. M. Fox, M. E. Raichle, and R. M. Burde. 1985. "The role of cerebral cortex in the generation of voluntary saccades: a positron emission tomographic study." *Journal of Neurophysiology* 54 (2): 348–369.
- Freeman, J., and E. P. Simoncelli. 2011. "Metamers of the ventral stream." *Nature Neuroscience* 14 (9): 1195–1201.
- French, A. N., R. S. Ashby, I. G. Morgan, and K. A. Rose. 2013. "Time outdoors and the prevention of myopia." *Experimental Eye Research* 114:58–68.
- Fresina, M., and E. C. Campos. 2014. "A 1-Year Review of Amblyopia and Strabismus Research." *The Asia-Pacific Journal of Ophthalmology* 3 (6): 379–387.
- Galvin, S. J., R. P. O'Shea, A. M. Squire, and D. G. Govan. 1997. "Sharpness overconstancy in peripheral vision." *Vision Research* 37 (15): 2035–2039.
- Galvin, S. J., R. P. O'Shea, A. M. Squire, and D. S. Hailstone. 1999. "Sharpness overconstancy: the roles of visibility and current context." *Vision Research* 39 (16): 2649–2657.
- Gauthier, G., J.-L. Vercher, F. M. Ivaldi, and E. Marchetti. 1988. "Oculo-manual tracking of visual targets: control learning, coordination control and coordination model." *Experimental Brain Research* 73 (1): 127–137.
- Gegenfurtner, K. R., M. Bloj, and M. Toscani. 2015. "The many colours of 'the dress'." *Current Biology* 25 (13): R543–R544.

- Geisler, W. S. 2008. “Visual perception and the statistical properties of natural scenes.” *Annual Review of Psychology* 59:167–192.
- Geisler, W. S., and J. S. Perry. 2002. “Real-time simulation of arbitrary visual fields.” In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, 83–87. ACM.
- Gibaldi, A., A. Canessa, and S. Sabatini. 2015. “The Dimpled Horopter Explained by the Strategy of Binocular Fixation.” In *Perception*, 44:353–354.
- Gibaldi, A., A. Canessa, M. Chessa, F. Solari, and S. P. Sabatini. 2012. “A neural model for coordinated control of horizontal and vertical alignment of the eyes in three-dimensional space.” In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 955–960. IEEE.
- Gibaldi, A., A. Canessa, and S. Sabatini. 2016. “Active stereo fixation: developmental influence on the binocular visual system.” *Journal of Vision* 16 (12): 430–430.
- Gibaldi, A., M. Vanegas, P. J. Bex, and G. Maiello. 2016. “Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research.” *Behavior Research Methods*: 1–24.
- Goldschmidt, E., and N. Jacobsen. 2014. “Genetic and environmental effects on myopia development and progression.” *Eye* 28 (2): 126–133.

- Gonzalez, D. A., and E. Niechwiej-Szwedo. 2016. “The effects of monocular viewing on hand-eye coordination during sequential grasping and placing movements.” *Vision Research* 128:30–38.
- Gonzalez, R. C., and R. E. Woods. 2008. *Digital Image Processing, 3rd edition*. Prentice Hall.
- Goodale, M. A., and D. A. Westwood. 2004. “An evolving view of duplex vision: separate but interacting cortical pathways for perception and action.” *Current Opinion in Neurobiology* 14 (2): 203–211.
- Gordon, C. C., T. Churchill, C. E. Clauser, B. Bradtmiller, and J. T. McConville. 1989. *Anthropometric survey of US army personnel: methods and summary statistics 1988*. Technical report. DTIC Document.
- Gorini, A., C. S. Capideville, G. De Leo, F. Mantovani, and G. Riva. 2011. “The role of immersion and narrative in mediated presence: the virtual hospital experience.” *Cyberpsychology, Behavior, and Social Networking* 14 (3): 99–105.
- Goss, D. A. 1991. “Clinical accommodation and heterophoria findings preceding juvenile onset of myopia.” *Optometry & Vision Science* 68 (2): 110–116.
- Grant, S., C. Suttle, D. R. Melmoth, M. L. Conway, and J. J. Sloper. 2014. “Age- and Stereovision-Dependent Eye-Hand Coordination Deficits in Children With Amblyopia and Abnormal Binocularity.” *Investigative Ophthalmology & Visual Science* 55 (9): 5687.

- Graziano, M., R. A. Andersen, and R. J. Snowden. 1994. "Tuning of MST neurons to spiral motions." *The Journal of Neuroscience* 14 (1): 54–67.
- Grealy, M. A., D. A. Johnson, and S. K. Rushton. 1999. "Improving cognitive function after brain injury: the use of exercise and virtual reality." *Archives of Physical Medicine and Rehabilitation* 80 (6): 661–667.
- Grossberg, S., E. Mingolla, and C. Pack. 1999. "A neural model of motion processing and visual navigation by cortical area MST." *Cerebral Cortex* 9 (8): 878–895.
- Grosvenor, T., and D. A. Goss. 1998. "Role of the cornea in emmetropia and myopia." *Optometry & Vision Science* 75 (2): 132–145.
- Grosvenor, T., and R. Scott. 1994. "Role of the axial length/corneal radius ratio in determining the refractive state of the eye." *Optometry & Vision Science* 71 (9): 573–579.
- Gu, Y., G. C. DeAngelis, and D. E. Angelaki. 2012. "Causal links between dorsal medial superior temporal area neurons and multisensory heading perception." *The Journal of Neuroscience* 32 (7): 2299–2313.
- Gu, Y., C. R. Fetsch, B. Adeyemo, G. C. DeAngelis, and D. E. Angelaki. 2010. "Decoding of MSTd population activity accounts for variations in the precision of heading perception." *Neuron* 66 (4): 596–609.
- Guna, J., G. Jakus, M. Pogačnik, S. Tomažič, and J. Sodnik. 2014. "An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking." *Sensors* 14 (2): 3702–3720.

- Gwiazda, J., J. Bauer, F. Thorn, and R. Held. 1995. "A dynamic relationship between myopia and blur-driven accommodation in school-aged children." *Vision Research* 35 (9): 1299–1304.
- Gwiazda, J., L. Hyman, T. T. Norton, M. E. Hussein, W. Marsh-Tootle, R. Manny, Y. Wang, and D. Everett. 2004. "Accommodation and related risk factors associated with myopia progression and their interaction with treatment in COMET children." *Investigative Ophthalmology & Visual Science* 45 (7): 2143–2151.
- Gwiazda, J., F. Thorn, J. Bauer, and R. Held. 1993. "Myopic children show insufficient accommodative response to blur." *Investigative Ophthalmology & Visual Science* 34 (3): 690–694.
- Gwiazda, J., F. Thorn, and R. Held. 2005. "Accommodation, accommodative convergence, and response AC/A ratios before and at the onset of myopia in children." *Optometry & Vision Science* 82 (4): 273–278.
- Hainline, L., and P. M. Riddell. 1995. "Binocular alignment and vergence in early infancy." *Vision Research* 35 (23): 3229–3236.
- Hainline, L., J. Turkel, I. Abramov, E. Lemerise, and C. M. Harris. 1984. "Characteristics of saccades in human infants." *Vision Research* 24 (12): 1771–1780.
- Hamerly, J. R., and C. A. Dvorak. 1981. "Detection and discrimination of blur in edges and lines." *Journal of the Optical Society of America* 71 (4): 448–452.

- Hampson, K. M., M. P. Cufflin, and E. A. Mallen. 2013. "Effect of correction of aberration dynamics on chaos in human ocular accommodation." *Optics Letters* 38 (22): 4747–4749.
- Hands, P., T. V. Smulders, and J. C. Read. 2015. "Stereoscopic 3-D content appears relatively veridical when viewed from an oblique angle." *Journal of Vision* 15 (5): 6.
- Harb, E., F. Thorn, and D. Troilo. 2006. "Characteristics of accommodative behavior during sustained reading in emmetropes and myopes." *Vision Research* 46 (16): 2581–2592.
- Haynes, H., B. L. White, and R. Held. 1965. "Visual accommodation in human infants." *Science* 148 (3669): 528–530.
- Hays, R. D., C. M. Mangione, L. Ellwein, A. S. Lindblad, K. L. Spritzer, P. J. McDonnell, N.-R. R. Group, et al. 2003. "Psychometric properties of the National Eye Institute–refractive error quality of life instrument." *Ophthalmology* 110 (12): 2292–2301.
- Heeger, D. J. 1992. "Normalization of cell responses in cat striate cortex." *Visual Neuroscience* 9 (2): 181–197.
- Heeger, D. J., E. P. Simoncelli, and J. A. Movshon. 1996. "Computational models of cortical visual processing." *Proceedings of the National Academy of Sciences* 93 (2): 623–627.
- Held, R. T., E. A. Cooper, and M. S. Banks. 2012. "Blur and Disparity Are Complementary Cues to Depth." *Current Biology* 22 (5): 426–431.

- Held, R. T., and M. S. Banks. 2008. "Misperceptions in stereoscopic displays: a vision science perspective." In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, 23–32. ACM.
- Held, R. T., E. A. Cooper, J. F. O'Brien, and M. S. Banks. 2010. "Using blur to affect perceived distance and size." *ACM Transactions on Graphics* 29 (2).
- Heravian, J., T. Jenkins, and W. Douthwaite. 1990. "Binocular summation in visually evoked responses and visual acuity." *Ophthalmic and Physiological Optics* 10 (3): 257–261.
- Hering, E. 1868. *Die Lehre vom binocularen Sehen*. Engelmann.
- Hess, R., B. Mansouri, and B. Thompson. 2010. "A new binocular approach to the treatment of amblyopia in adults well beyond the critical period of visual development." *Restorative Neurology and Neuroscience* 28 (6): 793–802.
- Hess, R. F., K. L. Schmid, S. O. Dumoulin, D. J. Field, and D. R. Brinkworth. 2006. "What image properties regulate eye growth?" *Current Biology* 16 (7): 687–691.
- Hikosaka, O., Y. Takikawa, and R. Kawagoe. 2000. "Role of the Basal Ganglia in the Control of Purposive Saccadic Eye Movements." *Physiological Reviews* 80 (3): 953–978.
- Hilz, R., and C. Cavonius. 1974. "Functional organization of the peripheral retina: sensitivity to periodic stimuli." *Vision Research* 14 (12): 1333–1337.



- Hoffman, D. M., and M. S. Banks. 2010. "Focus information is used to interpret binocular images." *Journal of Vision* 10 (5): 13.
- Hoffman, D. M., A. R. Girshick, K. Akeley, and M. S. Banks. 2008. "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue." *Journal of Vision* 8 (3): 33.
- Hoffman, H. G., W. J. Meyer III, M. Ramirez, L. Roberts, E. J. Seibel, B. Atzori, S. R. Sharar, and D. R. Patterson. 2014. "Feasibility of articulated arm mounted Oculus Rift Virtual Reality goggles for adjunctive pain control during occupational therapy in pediatric burn patients." *Cyberpsychology, Behavior, and Social Networking* 17 (6): 397–401.
- Hopp, J. J., and A. F. Fuchs. 2004. "The characteristics and neuronal substrate of saccadic eye movement plasticity." *Progress in Neurobiology* 72 (1): 27–53.
- Horwood, A. M., and P. M. Riddell. 2009. "Receding and disparity cues aid relaxation of accommodation." *Optometry and vision science: official publication of the American Academy of Optometry* 86 (11): 1276.
- Hou, F., L. Lesmes, P. Bex, M. Dorr, and Z.-L. Lu. 2015. "Using 10AFC to further improve the efficiency of the quick CSF method." *Journal of Vision* 15 (9): 2.
- Howe, C. Q., and D. Purves. 2002. "Range image statistics can explain the anomalous perception of length." *Proceedings of the National Academy of Sciences* 99 (20): 13184–13188.

- Huang, J., A. B. Lee, and D. B. Mumford. 2000. "Statistics of range images." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 324–331.
- Huang, J., L.-F. Hung, and E. L. Smith III. 2011. "Effects of foveal ablation on the pattern of peripheral refractive errors in normal and form-deprived infant rhesus monkeys (*Macaca mulatta*).” *Investigative Ophthalmology & Visual Science* 52 (9): 6428–6434.
- . 2012. "Recovery of peripheral refractive errors and ocular shape in rhesus monkeys (*Macaca mulatta*) with experimentally induced myopia.” *Vision Research* 73:30–39.
- Hung, G., K. Ciuffreda, J. Semmlow, and J. Horng. 1994. "Vergence eye movements under natural viewing conditions.” *Investigative Ophthalmology & Visual Science* 35 (9): 3486–3492.
- Hyams, S., E. Pokotilo, and G. Shkurko. 1977. "Prevalence of refractive errors in adults over 40: a survey of 8102 eyes.” *British Journal of Ophthalmology* 61 (6): 428–432.
- Ishio, H., T. Yamakawa, A. Sugiura, K. Yoshikawa, T. Kojima, S. Terada, K. Tanaka, and M. Miyao. 2015. "A Study on Within-Subject Factors for Visually Induced Motion Sickness by Using 8K Display.” In *International Conference on Universal Access in Human-Computer Interaction*, 196–204. Springer.

- Jacobs, R., G. Smith, and C. Chan. 1989. "Effect of defocus on blur thresholds and on thresholds of perceived change in blur: comparison of source and observer methods." *Optometry & Vision Science* 66 (8): 545–553.
- Jansen, L., S. Onat, and P. König. 2009. "Influence of disparity on fixation and saccades in free viewing of natural scenes." *Journal of Vision* 9 (1): 29.
- Javal, E. 1879. "Essai sur la physiologie de la lecture." In *Annales D'Oculistique*.
- Jiang, B.-C. 1997. "Integration of a sensory component into the accommodation model reveals differences between emmetropia and late-onset myopia." *Investigative Ophthalmology & Visual Science* 38 (8): 1511–1516.
- Johnson, P. V., J. A. Parnell, J. Kim, C. D. Saunter, G. D. Love, and M. S. Banks. 2016. "Dynamic lens and monovision 3D displays to improve viewer comfort." *Optics Express* 24 (11): 11808–11827.
- Johnston, E. B., B. G. Cumming, and A. J. Parker. 1993. "Integration of depth modules: Stereopsis and texture." *Vision Research* 33 (5): 813–826.
- Johnston, I. R., G. R. White, and R. W. Cumming. 1973. "The role of optical expansion patterns in locomotor control." *The American Journal of Psychology*: 311–324.
- Jones, L. A., L. T. Sinnott, D. O. Mutti, G. L. Mitchell, M. L. Moeschberger, and K. Zadnik. 2007. "Parental history of myopia, sports and outdoor activities, and future myopia." *Investigative Ophthalmology & Visual Science* 48 (8): 3524–3532.

- Julesz, B. 1971. *Foundations of Cyclopean Perception*. Chicago: University of Chicago press.
- Kamiya, S. 1984. "Analysis of school myopia using the Nidek Auto-Refractometer AR 3000. Part 2. Observation of frequency distribution of pupil refraction." *Folia Ophthalmologica Japonica* 35:1755–1769.
- Kasthurirangan, S., and A. Glasser. 2005a. "Characteristics of pupil responses during far-to-near and near-to-far accommodation." *Ophthalmic and Physiological Optics* 25 (4): 328–339.
- . 2005b. "Influence of amplitude and starting point on accommodative dynamics in humans." *Investigative Ophthalmology & Visual Science* 46 (9): 3463–3472.
- Kasthurirangan, S., A. S. Vilupuru, and A. Glasser. 2003. "Amplitude dependent accommodative dynamics in humans." *Vision Research* 43 (27): 2945–2956.
- Katz, S. D. 1991. *Film directing shot by shot: visualizing from concept to screen*. Gulf Professional Publishing.
- Kennedy, R. S., N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. 1993. "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness." *The International Journal of Aviation Psychology* 3 (3): 203–220.
- Kerber, K. L., F. Thorn, P. J. Bex, and F. A. Vera-Diaz. 2016. "Peripheral contrast sensitivity and attention in myopia." *Vision Research* 125:49–54.

- Kim, J., C. Y. Chung, S. Nakamura, S. Palmisano, and S. K. Khuu. 2015. "The Oculus Rift: a cost-effective tool for studying visual-vestibular interactions in self-motion perception." *Frontiers in Psychology* 6.
- King, W. M. 2011. "Binocular coordination of eye movements - Hering's Law of equal innervation or uniocular control?" *European Journal of Neuroscience* 33 (11): 2139–2146.
- King, W., and W. Zhou. 2002. "Neural basis of disjunctive eye movements." *Annals of the New York Academy of Sciences* 956 (1): 273–283.
- Knox, P. J., A. J. Simmers, L. S. Gray, and M. Cleary. 2012. "An exploratory study: prolonged periods of binocular stimulation can provide an effective treatment for childhood amblyopia." *Investigative Ophthalmology & Visual Science* 53 (2): 817–824.
- Koenderink, J. J. 1986. "Optic flow." *Vision Research* 26 (1): 161–179.
- Koken, P. W., and C. J. Erkelens. 1992. "Influences of hand movements on eye movements in tracking tasks in man." *Experimental Brain Research* 88 (3): 657–664.
- . 1993. "Simultaneous hand tracking does not affect human vergence pursuit." *Experimental Brain Research* 96 (3): 494–500.
- Kooi, F. L., and A. Toet. 2004. "Visual comfort of binocular and 3D displays." *Displays* 25 (2): 99–108.
- Kosara, R., S. Miksch, and H. Hauser. 2001. "Semantic depth of field." In *infovvis*, 97. IEEE.

- Kowler, E. 2011. "Eye movements: The past 25years." *Vision Research* 51 (13): 1457–1483.
- Krauskopf, J., T. Cornsweet, and L. Riggs. 1960. "Analysis of eye movements during monocular and binocular fixation." *Journal of the Optical Society of America*.
- Kretzmer, E. 1952. "Statistics of television signals." *Bell System Technical Journal* 31 (4): 751–763.
- Krijn, M., P. M. Emmelkamp, R. P. Olafsson, and R. Biemond. 2004. "Virtual reality exposure therapy of anxiety disorders: A review." *Clinical Psychology Review* 24 (3): 259–281.
- Kwon, J. H., J. Powell, and A. Chalmers. 2013. "How level of realism influences anxiety in virtual reality environments for a job interview." *International Journal of Human-Computer Studies* 71 (10): 978–987.
- Lafer-Sousa, R., K. L. Hermann, and B. R. Conway. 2015. "Striking individual differences in color perception uncovered by 'the dress' photograph." *Current Biology* 25 (13): R545–R546.
- Lambooij, M., M. Fortuin, I. Heynderickx, and W. IJsselstein. 2009. "Visual discomfort and visual fatigue of stereoscopic displays: a review." *Journal of Imaging Science and Technology* 53 (3): 30201–1.
- Landy, M. S., L. T. Maloney, E. B. Johnston, and M. Young. 1995. "Measurement and modeling of depth cue combination: In defense of weak fusion." *Vision Research* 35 (3): 389–412.

- Langaas, T., P. M. Riddell, E. Svarverud, A. E. Ystenæs, I. Langeeggen, and J. R. Bruenech. 2008. "Variability of the accommodation response in early onset myopia." *Optometry & Vision Science* 85 (1): 37–48.
- Langer, M. S., and R. A. Siciliano. 2015. "Are blur and disparity complementary cues to depth?" *Vision Research* 107:15–21.
- Lappe, M., and J. P. Rauschecker. 1993. "A neural network for the processing of optic flow from ego-motion in man and higher mammals." *Neural Computation* 5 (3): 374–391.
- Laver, K., S. George, S. Thomas, J. E. Deutsch, and M. Crotty. 2012. "Virtual reality for stroke rehabilitation." *Stroke* 43 (2): e20–e21.
- Lazzari, S., J.-L. Vercher, and A. Buizza. 1997. "Manuo-ocular coordination in target tracking. I. A model simulating human performance." *Biological Cybernetics* 77 (4): 257–266.
- Lebreton, P., A. Raake, M. Barkowsky, and P. Le Callet. 2014. "Measuring perceived depth in natural images and study of its relation with monocular and binocular depth cues." In *IS&T/SPIE Electronic Imaging*, 90110C–90110C. International Society for Optics and Photonics.
- Lee, A. B., D. Mumford, and J. Huang. 2001. "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model." *International Journal of Computer Vision* 41 (1-2): 35–59.
- Legge, G. E. 1984a. "Binocular contrast summation-I. Detection and discrimination." *Vision Research* 24 (4): 373–383.

- Legge, G. E. 1984b. "Binocular contrast summation-II. Quadratic summation." *Vision Research* 24 (4): 385–394.
- Lemij, H., and H. Collewijn. 1991a. "Long-term nonconjugate adaptation of human saccades to anisometropic spectacles." *Vision Research* 31 (11): 1939–1954.
- . 1991b. "Short-term nonconjugate adaptation of human saccades to anisometropic spectacles." *Vision Research* 31 (11): 1955–1966.
- Lesmes, L. A., Z.-L. Lu, J. Baek, and T. D. Albright. 2010. "Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method." *Journal of Vision* 10 (3): 17–17.
- Lessiter, J., J. Freeman, E. Keogh, and J. Davidoff. 2001. "A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory." *Presence: Teleoperators and Virtual Environments* 10 (3): 282–297.
- Leyland, M., and E. Zinicola. 2003. "Multifocal versus monofocal intraocular lenses in cataract surgery: a systematic review." *Ophthalmology* 110 (9): 1789–1798.
- Li, W.-J., C.-C. Chang, K.-Y. Hsu, M.-D. Kuo, and D.-L. Way. 2001. "A PC-based distributed multiple display virtual reality system." *Displays* 22 (5): 177–181.
- Li, J., B. Thompson, D. Deng, L. Y. Chan, M. Yu, and R. F. Hess. 2013. "Dichoptic training enables the adult amblyopic brain to learn." *Current Biology* 23 (8): R308–R309.



- Lin, L., Y. Shih, C. Hsiao, and C. Chen. 2004. "Prevalence of myopia in Taiwanese schoolchildren: 1983 to 2000." *Annals-Academy of Medicine Singapore* 33 (1): 27–33.
- Lin, Z., B. Vasudevan, Y. B. Liang, Y. C. Zhang, S. Q. Zhao, X. D. Yang, N. L. Wang, B. Gilmartin, and K. J. Ciuffreda. 2013. "Nearwork-induced transient myopia (NITM) in anisometropia." *Ophthalmic and Physiological Optics* 33 (3): 311–317.
- Lisberger, S. G. 2015. "Visual Guidance of Smooth Pursuit Eye Movements." *Annual Review of Vision Science* 1:447–468.
- Liu, S., and H. Hua. 2010. "A systematic method for designing depth-fused multi-focal plane three-dimensional displays." *Optics Express* 18 (11): 11562–11573.
- Liu, Y., A. C. Bovik, and L. K. Cormack. 2008. "Disparity statistics in natural scenes." *Journal of Vision* 8 (11): 19.
- López-Gil, N., J. Martin, T. Liu, A. Bradley, D. Díaz-Muñoz, and L. N. Thi-bos. 2013. "Retinal image quality during accommodation." *Ophthalmic and Physiological Optics* 33 (4): 497–507.
- Loschky, L., and G. Wolverton. 2007. "How late can you update gaze-contingent multiresolutional displays without detection?" *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)* 3 (4): 7.

- Loschky, L. C., and G. W. McConkie. 2002. “Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display.” *Journal of Experimental Psychology: Applied* 8 (2): 99.
- Love, G. D., D. M. Hoffman, P. J. Hands, J. Gao, A. K. Kirby, and M. S. Banks. 2009. “High-speed switchable lens enables the development of a volumetric stereoscopic display.” *Optics Express* 17 (18): 15716–15725.
- Lutgheid, A. J., L. M. Wilcox, R. S. Allison, and I. P. Howard. 2014. “Vergence eye movements are not essential for stereoscopic depth.” *Proceedings of the Royal Society B: Biological Sciences* 281 (1776): 2013–2118.
- MacKenzie, K. J., D. M. Hoffman, and S. J. Watt. 2010. “Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control.” *Journal of Vision* 10 (8): 22–22.
- Maiello, G., M. Chessa, F. Solari, and P. J. Bex. 2014. “Simulated disparity and peripheral blur interact during binocular fusion.” *Journal of Vision* 14 (8): 13.
- . 2015a. “A Space-Variant Model for Motion Interpretation across the Visual Field.” In *MODVIS*.
- Maiello, G., P. J. Bex, and F. A. Vera-Diaz. 2014. *Prevention and Treatment of Myopia*. US Patent App. 14/913,586.
- Maiello, G., M. Chessa, P. J. Bex, and F. Solari. 2016. “Can Neuromorphic Computer Vision Inform Vision Science? Disparity Estimation as a Case Study.” In *MODVIS*.

- Maiello, G., M. Chessa, F. Solari, and P. J. Bex. 2015b. “The (In) Effectiveness of Simulated Blur for Depth Perception in Naturalistic Images.” *PloS one* 10 (10): e0140230.
- Maiello, G., A. Gibaldi, S. S. P, and P. J. Bex. 2016. “Vergence Eye Movements to Unbalanced Dichoptic Visual Stimuli.” In *European Conference on Visual Perception*.
- Maiello, G., W. Harrison, F. Vera-Diaz, and P. J. Bex. 2015. “Perceptual Consequences of Elongated Eyes.” *Journal of Vision* 15 (12): 111–111.
- Marr, D., and T. Poggio. 1979. “A Computational Theory of Human Stereo Vision.” *Proceedings of the Royal Society of London. Series B, Biological Sciences* 204 (1156): 301–328.
- Marshall, J. A., C. A. Burbeck, D. Ariely, J. P. Rolland, and K. E. Martin. 1996. “Occlusion edge blur: a cue to relative visual depth.” *Journal of the Optical Society of America A* 13 (4): 681–688.
- Mather, G. 1996. “Image Blur as a Pictorial Depth Cue.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 263 (1367): 169–172.
- Mather, G., and D. R. Smith. 2000. “Depth cue integration: stereopsis and image blur.” *Vision Research* 40 (25): 3501–3506.
- Mather, G. 1997. “The use of image blur as a depth cue.” *Perception* 26 (9): 1147–1158.
- Mather, G., and D. R. Smith. 2002. “Blur discrimination and its relation to blur-mediated depth perception.” *Perception* 31 (10): 1211–1219.

- Mather, J. A., and J. R. Lackner. 1980. "Visual tracking of active and passive movements of the hand." *Quarterly Journal of Experimental Psychology* 32 (2): 307–315.
- . 1981. "The influence of efferent, proprioceptive, and timing factors on the accuracy of eye-hand tracking." *Experimental Brain Research* 43 (3-4): 406–412.
- Mauderer, M., S. Conte, M. A. Nacenta, and D. Vishwanath. 2014. "Depth perception with gaze-contingent depth of field." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 217–226. ACM.
- Maxwell, J. S., and W. M. King. 1992. "Dynamics and efficacy of saccade-facilitated vergence eye movements in monkeys." *Journal of Neurophysiology* 68 (4): 1248–1260.
- McAnany, J. J., and K. R. Alexander. 2006. "Contrast sensitivity for letter optotypes vs. gratings under conditions biased toward parvocellular and magnocellular pathways." *Vision Research* 46 (10): 1574–1584.
- McBrient, N. A., and M. Millodot. 1986. "The effect of refractive error on the accommodative response gradient." *Ophthalmic and Physiological Optics* 6 (2): 145–149.
- McKee, S. P., and K. Nakayama. 1984. "The detection of motion in the peripheral visual field." *Vision Research* 24 (1): 25–32.
- McLaughlin, S. C. 1967. "Parametric adjustment in saccadic eye movements." *Perception & Psychophysics* 2 (8): 359–362.

- Metta, G., G. Sandini, D. Vernon, L. Natale, and F. Nori. 2008. "The iCub humanoid robot: an open platform for research in embodied cognition." In *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, 50–56. ACM.
- Miller Jr, R. G. 2011. *Survival analysis*. Vol. 66. John Wiley & Sons.
- Millodot, M. 2015. "The effect of refractive error on the accommodative response gradient: a summary and update." *Ophthalmic and Physiological Optics* 35 (6): 607–612.
- Morgan, I. G., K. Ohno-Matsui, and S.-M. Saw. 2012. "Myopia." *The Lancet* 379 (9827): 1739–1748.
- Moscatelli, A., M. Mezzetti, and F. Lacquaniti. 2012. "Modeling psychophysical data at the population-level: the generalized linear mixed model." *Journal of Vision* 12 (11): 26.
- Murray, S., and P. J. Bex. 2010. "Perceived blur in naturally contoured images depends on phase." *Frontiers in Psychology* 1:185.
- Mutti, D. O., J. R. Hayes, G. L. Mitchell, L. A. Jones, M. L. Moeschberger, S. A. Cotter, R. N. Kleinstein, R. E. Manny, J. D. Twelker, and K. Zadnik. 2007. "Refractive error, axial length, and relative peripheral refractive error before and after the onset of myopia." *Investigative Ophthalmology & Visual Science* 48 (6): 2510–2519.
- Nefs, H. T. 2012. "Depth of field affects perceived depth-width ratios in photographs of natural scenes." *Seeing and Perceiving* 25 (6): 577–595.

- Ng, R. 2006. “Digital light field photography.” PhD diss., Stanford University.
- Ng, R., M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. 2005. “Light field photography with a hand-held plenoptic camera.” *Computer Science Technical Report CSTR 2* (11).
- Ngo, C., S.-M. Saw, R. Dharani, and I. Flitcroft. 2013. “Does sunlight (bright lights) explain the protective effects of outdoor activity against myopia?” *Ophthalmic and Physiological Optics* 33 (3): 368–372.
- Nguyen, V. A., I. P. Howard, and R. S. Allison. 2005. “Detection of the depth order of defocused images.” *Vision Research* 45 (8): 1003–1011.
- Niehorster, D. C., W. W. F. Siu, and L. Li. 2015. “Manual tracking enhances smooth pursuit eye movements.” *Journal of Vision* 15 (15): 11.
- Noorden, G. K. von, and G. Mackensen. 1962. “Pursuit Movements of Normal and Amblyopic Eyes: An Electro-Ophthalmographic Study 1. Physiology of Pursuit Movements.” *American Journal of Ophthalmology* 53 (2): 325–336.
- North, M. M., S. M. North, and J. R. Coble. 1996. “Effectiveness of virtual environment desensitization in the treatment of agoraphobia.” *Presence: Teleoperators & Virtual Environments* 5 (3): 346–352.
- Norton, T. T., J. T. Siegwart, and A. O. Amedo. 2006. “Effectiveness of hyperopic defocus, minimal defocus, or myopic defocus in competition with a myopiagenic stimulus in tree shrew eyes.” *Investigative Ophthalmology & Visual Science* 47 (11): 4687–4699.

- O'Brien, J., and A. Johnston. 2000. "When texture takes precedence over motion in depth perception." *Perception* 29 (4): 437–452.
- O'Connor, A. R., E. E. Birch, S. Anderson, and H. Draper. 2010. "The functional significance of stereopsis." *Investigative Ophthalmology & Visual Science* 51 (4): 2019–2023.
- O'Driscoll, G. A., N. M. Alpert, S. W. Matthysse, D. L. Levy, S. L. Rauch, and P. S. Holzman. 1995. "Functional neuroanatomy of antisaccade eye movements investigated with positron emission tomography." *Proceedings of the National Academy of Sciences* 92 (3): 925–929.
- Ohzawa, I., G. C. DeAngelis, R. D. Freeman, et al. 1990. "Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors." *Science* 249 (4972): 1037–1041.
- Ooms, K., L. Dupont, L. Lapon, and S. Popelka. 2015. "Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental set-ups." *Journal of Eye Movement Research* 8 (1): 1–24.
- Orban, G., L. Lagae, A. Verri, S. Raiguel, D. Xiao, H. Maes, and V. Torre. 1992. "First-order analysis of optical flow in monkey brain." *Proceedings of the National Academy of Sciences* 89 (7): 2595–2599.
- Orban, G. A. 2008. "Higher order visual processing in macaque extrastriate cortex." *Physiological Reviews* 88 (1): 59–89.

- Oshika, T., C. Okamoto, T. Samejima, T. Tokunaga, and K. Miyata. 2006. "Contrast sensitivity function and ocular higher-order wavefront aberrations in normal human eyes." *Ophthalmology* 113 (10): 1807–1812.
- Oshima, S. 1958. "Studies on the depth-of-focus of the eye." *Japanese Journal of Ophthalmology* 2:63–72.
- Pääkkönen, A. K., and M. J. Morgan. 1994. "Effects of motion on blur discrimination." *Journal of the Optical Society of America A* 11 (3): 992–1002.
- Pamplona, D., and A. Bernardino. 2009. "Smooth foveal vision with Gaussian receptive fields." In *2009 9th IEEE-RAS International Conference on Humanoid Robots*, 223–229. IEEE.
- Pan, C.-W., D. Ramamurthy, and S.-M. Saw. 2012. "Worldwide prevalence and risk factors for myopia." *Ophthalmic and Physiological Optics* 32 (1): 3–16.
- Pandian, A., P. R. Sankaridurg, T. Naduvilath, D. O’Leary, D. F. Sweeney, K. Rose, and P. Mitchell. 2006. "Accommodative facility in eyes with and without myopia." *Investigative Ophthalmology & Visual Science* 47 (11): 4725–4731.
- Patel, N. 2012. *Reverse engineering the lytro .lfp file format*. Retrieved online: <http://eclecti.cc/computervision/reverse-engineering-the-lytro-lfp-file-format>.



- Péllisson, D., N. Alahyane, M. Panouilleres, and C. Tilikete. 2010. "Sensorimotor adaptation of saccadic eye movements." *Neuroscience & Biobehavioral Reviews* 34 (8): 1103–1120.
- Péllisson, D., L. Goffart, A. Guillaume, N. Catz, and G. Raboyeau. 2001. "Early head movements elicited by visual stimuli or collicular electrical stimulation in the cat." *Vision Research* 41 (25): 3283–3294.
- Pelli, D. G. 1997. "The VideoToolbox software for visual psychophysics: Transforming numbers into movies." *Spatial Vision* 10:437–442.
- Pelli, D. G., and P. J. Bex. 2013. "Measuring contrast sensitivity." *Vision Research* 90:10–14.
- Pelli, D. 1990. "Vision: Coding and efficiency." Chap. The quantum efficiency of vision, edited by C. Blakemore, 3–24. Cambridge: Cambridge University Press.
- Perrone, J. A., and L. S. Stone. 1994. "A model of self-motion estimation within primate extrastriate visual cortex." *Vision Research* 34 (21): 2917–2938.
- Perry, C. J., P. Amarasekariya, and M. Fallah. 2016. "An eye in the palm of your hand: alterations in visual processing near the hand, a mini-review." *Frontiers in Computational Neuroscience* 10.
- Phillips, L., V. Interrante, M. Kaeding, B. Ries, and L. Anderson. 2012. "Correlations between physiological response, gait, personality, and presence in immersive virtual environments." *Presence: Teleoperators and Virtual Environments* 21 (2): 119–141.

- Plainis, S., D. Petratos, T. Giannakopoulou, D. A. Atchison, and M. K. Tsilimbaris. 2011. “Binocular summation improves performance to defocus-induced blur.” *Investigative Ophthalmology & Visual Science* 52 (5): 2784–2789.
- Pouget, A., K. Zhang, S. Deneve, and P. E. Latham. 1998. “Statistically efficient estimation using population coding.” *Neural Computation* 10 (2): 373–401.
- Poulere, E., J. Moschandreas, G. A. Kontadakis, I. G. Pallikaris, and S. Plainis. 2013. “Effect of blur and subsequent adaptation on visual acuity using letter and Landolt C charts: differences between emmetropes and myopes.” *Ophthalmic and Physiological Optics* 33 (2): 130–137.
- Price, H., P. M. Allen, H. Radhakrishnan, R. Calver, S. Rae, B. Theagarayan, A. Sailoganathan, and D. J. O’Leary. 2013. “The Cambridge Anti-myopia Study: variables associated with myopia progression.” *Optometry & Vision Science* 90 (11): 1274–1283.
- Rad, K. R., and L. Paninski. 2011. “Information Rates and Optimal Decoding in Large Neural Populations.” In *NIPS*, edited by J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, 846–854.
- Ramat, S., V. E. Das, J. T. Somers, and R. J. Leigh. 1999. “Tests of two hypotheses to account for different-sized saccades during disjunctive gaze shifts.” *Experimental Brain Research* 129 (4): 500–510.
- Rasch, G. 1960. *Probabilistic models for some intelligence and attainment tests*. Danish National Institute for Educational Research.

- Repetto, C., and G. Riva. 2011. "From virtual reality to interreality in the treatment of anxiety disorders." *Neuropsychiatry* 1 (1): 31–43.
- Resnikoff, S., D. Pascolini, S. P. Mariotti, and G. P. Pokharel. 2008. "Global magnitude of visual impairment caused by uncorrected refractive errors in 2004." *Bulletin of the World Health Organization* 86 (1): 63–70.
- Riddell, P. M., L. Hainline, and I. Abramov. 1994. "Calibration of the Hirschberg test in human infants." *Investigative Ophthalmology & Visual Science* 35 (2): 538–543.
- Rideout, V. J., U. G. Foehr, and D. F. Roberts. 2011. *Generation M2: Media in the Lives of 8 to 18 year Olds: A Kaiser Family Foundation Study, January 2010*.
- Robinson, D. 1965. "The mechanics of human smooth pursuit eye movement." *The Journal of Physiology* 180 (3): 569–591.
- Robinson, F. R., and A. F. Fuchs. 2001. "The role of the cerebellum in voluntary eye movements." *Annual Review of Neuroscience* 24 (1): 981–1004.
- Rolfs, M. 2009. "Microsaccades: small steps on a long way." *Vision Research* 49 (20): 2415–2441.
- Ronchi, L., and G. Molesini. 1975. "Depth of focus in peripheral vision." *Ophthalmic Research* 7 (3): 152–157.

- Roorda, A., M. C. Campbell, and W. R. Bobier. 1997. "Slope-based eccentric photorefraction: theoretical analysis of different light source configurations and effects of ocular aberrations." *Journal of the Optical Society of America A* 14 (10): 2547–2556.
- Rose, K. A., I. G. Morgan, J. Ip, A. Kifley, S. Huynh, W. Smith, and P. Mitchell. 2008. "Outdoor activity reduces the prevalence of myopia in children." *Ophthalmology* 115 (8): 1279–1285.
- Rosén, R., L. Lundström, and P. Unsbo. 2011. "Influence of optical defocus on peripheral vision." *Investigative Ophthalmology & Visual Science* 52 (1): 318–323.
- Rosenfield, M., and J. A. Abraham-Cohen. 1999. "Blur sensitivity in myopes." *Optometry & Vision Science* 76 (5): 303–307.
- Rosenfield, M., R. Desai, J. K. Portello, et al. 2002. "Do progressing myopes show reduced accommodative responses?" *Optometry & Vision Science* 79 (4): 268–273.
- Ross, J., M. Morrone, M. Goldberg, and D. Burr. 2001. "Changes in visual perception at the time of saccades." *Trends in Neurosciences* 24 (2): 113–121.
- Roucoux, A., C. Culee, and M. Roucoux. 1983. "Development of fixation and pursuit eye movements in human infants." *Behavioural Brain Research* 10 (1): 133–139.

- Rubin, G. S., B. Munoz, K. Bandeen-Roche, and S. K. West. 2000. "Monocular versus binocular visual acuity as measures of vision impairment and predictors of visual disability." *Investigative Ophthalmology & Visual Science* 41 (11): 3327–3334.
- Rushton, S. K., J. M. Harris, M. R. Lloyd, and J. P. Wann. 1998. "Guidance of locomotion on foot uses perceived target location rather than optic flow." *Current Biology* 8 (21): 1191–1194.
- Rushton, S. K., and P. M. Riddell. 1999. "Developing visual systems and exposure to virtual reality and stereo displays: some concerns and speculations about the demands on accommodation and vergence." *Applied Ergonomics* 30 (1): 69–78.
- Rushton, S. K., and P. A. Warren. 2005. "Moving observers, relative retinal motion and the detection of object movement." *Current Biology* 15 (14): R542–R543.
- Rushton, S. K., J. Wen, and R. S. Allison. 2002. "Egocentric direction and the visual guidance of robot locomotion background, theory and implementation." In *International Workshop on Biologically Motivated Computer Vision*, 576–591.
- Rushton, S., M. Mon-Williams, and J. P. Wann. 1994. "Binocular vision in a bi-ocular world: new-generation head-mounted displays avoid causing visual deficit." *Displays* 15 (4): 255–260.
- Salvi, S. M., S. Akhtar, and Z. Currie. 2006. "Ageing changes in the eye." *Postgraduate Medical Journal* 82 (971): 581–587.

- Saslow, M. 1967. "Latency for saccadic eye movement." *Journal of the Optical Society of America* 57 (8): 1030–1033.
- Saunders, D. R., and R. L. Woods. 2014. "Direct measurement of the system latency of gaze-contingent displays." *Behavior Research Methods* 46 (2): 439–447.
- Saw, S.-M. 2003. "A synopsis of the prevalence rates and environmental risk factors for myopia." *Clinical and Experimental Optometry* 86 (5): 289–294.
- Saw, S.-M., G. Gazzard, K. A. Eong, and D. Tan. 2002. "Myopia: attempts to arrest progression." *British Journal of Ophthalmology* 86 (11): 1306–1311.
- Sawides, L., P. de Gracia, C. Dorronsoro, M. Webster, and S. Marcos. 2011. "Adapting to blur produced by ocular high-order aberrations." *Journal of Vision* 11 (7): 21.
- Scarchilli, K., and J.-L. Vercher. 1999. "The oculomanual coordination control center takes into account the mechanical properties of the arm." *Experimental Brain Research* 124 (1): 42–52.
- Schaeffel, F., H. Wilhelm, and E. Zrenner. 1993. "Inter-individual variability in the dynamics of natural accommodation in humans: relation to age and refractive errors." *The Journal of Physiology* 461:301.
- Schall, J. D., and K. G. Thompson. 1999. "Neural selection and control of visually guided eye movements." *Annual Review of Neuroscience* 22 (1): 241–259.

- Schenk, T., and R. D. McIntosh. 2010. "Do we have independent visual streams for perception and action?" *Cognitive Neuroscience* 1 (1): 52–62.
- Schmid, K. L., D. R. Iskander, R. W. Li, M. H. Edwards, and J. K. Lew. 2002. "Blur detection thresholds in childhood myopia: single and dual target presentation." *Vision Research* 42 (2): 239–247.
- Schmid, K. L., and N. C. Strang. 2015. "Differences in the accommodation stimulus response curves of adult myopes and emmetropes: a summary and update." *Ophthalmic and Physiological Optics* 35 (6): 613–621.
- Schor, C. M., and K. J. Ciuffreda. 1983. *Vergence eye Movements: Basic and Clinical Aspects*. Boston: Butterworth.
- Schor, C. M., I. Wood, and J. Ogawa. 1984a. "Binocular sensory fusion is limited by spatial resolution." *Vision Research* 24 (7): 661–665.
- Schor, C. M. 1999. "The influence of interactions between accommodation and convergence on the lag of accommodation." *Ophthalmic and Physiological Optics* 19 (2): 134–150.
- . 2011. "Adler's Physiology of the Eye." Chap. Neural control of eye movements, 10:830–858.
- Schor, C. M., I. Wood, and J. Ogawa. 1984b. "Binocular sensory fusion is limited by spatial resolution." *Vision Research* 24 (7): 661–665.
- Schultz, K. P., and C. Busettini. 2013. "Short-term saccadic adaptation in the macaque monkey: a binocular mechanism." *Journal of Neurophysiology* 109 (2): 518–545.

- Schumer, R. A., and B. Julesz. 1984. "Binocular disparity modulation sensitivity to disparities offset from the plane of fixation." *Vision Research* 24 (6): 533–541.
- Schwartz, E. 1977. "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception." *Biological Cybernetics* 25:181–194.
- Scott, R., and T. Grosvenor. 1993. "Structural model for emmetropic and myopic eyes." *Ophthalmic and Physiological Optics* 13 (1): 41–47.
- Sebastian, S., J. Burge, and W. S. Geisler. 2015. "Defocus blur discrimination in natural images with natural optics." *Journal of Vision* 15 (5): 16–16.
- Seidel, D., L. S. Gray, and G. Heron. 2005. "The effect of monocular and binocular viewing on the accommodation response to real targets in emmetropia and myopia." *Optometry & Vision Science* 82 (4): 279–285.
- Seidel, D., L. S. Gray, and G. Heron. 2003. "Retinotopic accommodation responses in myopia." *Investigative Ophthalmology & Visual Science* 44 (3): 1035–1041.
- Semmlow, J., and P. Wetzl. 1979. "Dynamic contributions of the components of binocular vergence." *Journal of the Optical Society of America* 69 (5): 639–645.
- Sharples, S., S. Cobb, A. Moody, and J. R. Wilson. 2008. "Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems." *Displays* 29 (2): 58–69.



- Shaw, R. B. J., E. B. Katzel, P. F. Koltz, M. J. Yaremchuk, J. A. Girotto, D. M. Kahn, and H. N. Langstein. 2011. "Aging of the facial skeleton: aesthetic implications and rejuvenation strategies." *Plastic and Reconstructive Surgery* 127 (1): 374–383.
- Sherman, S. M., T. Norton, and V. Casagrande. 1977. "Myopia in the lid-sutured tree shrew (*Tupaia glis*)." *Brain Research* 124 (1): 154–157.
- Shibata, T., J. Kim, D. M. Hoffman, and M. S. Banks. 2011. "The zone of comfort: Predicting visual discomfort with stereo displays." *Journal of Vision* 11 (8): 11–11.
- Shimono, K., H. Ono, S. Saida, and A. P. Mapp. 1998. "Methodological caveats for monitoring binocular eye position with nonius stimuli." *Vision Research* 38 (4): 591–600.
- Siderov, J., and R. S. Harwerth. 1995. "Stereopsis, spatial frequency and retinal eccentricity." *Vision Research* 35 (16): 2329–2337.
- Simoncelli, E. P., and D. J. Heeger. 1998. "A model of neuronal responses in visual area MT." *Vision Research* 38 (5): 743–761.
- Slater, M., C. Guger, G. Edlinger, R. Leeb, G. Pfurtscheller, A. Antley, M. Garau, A. Brogni, and D. Friedman. 2006. "Analysis of physiological responses to a social situation in an immersive virtual environment." *Presence: Teleoperators and Virtual Environments* 15 (5): 553–569.
- Slater, M., M. Usoh, and A. Steed. 1994. "Depth of presence in virtual environments." *Presence: Teleoperators & Virtual Environments* 3 (2): 130–144.

- Smith III, E. L. 2013. "Optical treatment strategies to slow myopia progression: effects of the visual extent of the optical treatment zone." *Experimental Eye Research* 114:77–88.
- Smith III, E. L., M. C. Campbell, and E. Irving. 2013. "Does peripheral retinal input explain the promising myopia control effects of corneal reshaping therapy (CRT or ortho-K) & multifocal soft contact lenses?" *Ophthalmic and Physiological Optics* 33 (3): 379–384.
- Smith III, E. L., H. Li-Fang, and R. S. Harwerth. 1994. "Effects of optically induced blur on the refractive status of young monkeys." *Vision Research* 34 (3): 293–301.
- Smith III, E. L., and L.-F. Hung. 1999. "The role of optical defocus in regulating refractive development in infant monkeys." *Vision Research* 39 (8): 1415–1435.
- Smith III, E. L., L.-F. Hung, and J. Huang. 2012. "Protective effects of high ambient lighting on the development of form-deprivation myopia in rhesus monkeys." *Investigative Ophthalmology & Visual Science* 53 (1): 421–428.
- Smith III, E. L., R. Ramamirtham, Y. Qiao-Grider, L.-F. Hung, J. Huang, C.-s. Kee, D. Coats, and E. Paysse. 2007. "Effects of foveal ablation on emmetropization and form-deprivation myopia." *Investigative Ophthalmology & Visual Science* 48 (9): 3914–3922.
- Smith, G. 1982. "Angular diameter of defocus blur discs." *Optometry & Vision Science* 59 (11): 885–889.

- Solari, F., M. Chessa, N. K. Medathati, and P. Kornprobst. 2015. "What can we expect from a V1-MT feedforward architecture for optical flow estimation?" *Signal Processing: Image Communication* 39:342–354.
- Solari, F., M. Chessa, and S. P. Sabatini. 2012. "Design strategies for direct multi-scale and multi-orientation feature extraction in the log-polar domain." *Pattern Recognition Letters* 33 (1): 41–51.
- . 2014. "An integrated neuromimetic architecture for direct motion interpretation in the log-polar domain." *Computer Vision and Image Understanding* 125:37–54.
- Solimini, A. G. 2013. "Are there side effects to watching 3D movies? A prospective crossover observational study on visually induced motion sickness." *PloS one* 8 (2): e56160.
- Sorsby, A., M. Sheridan, G. A. Leary, and B. Benjamin. 1960. "Vision, visual acuity, and ocular refraction of young men." *British Medical Journal* 1 (5183): 1394.
- Sparks, D. L. 1999. "Conceptual issues related to the role of the superior colliculus in the control of gaze." *Current Opinion in Neurobiology* 9 (6): 698–707.
- Spering, M., and A. Montagnini. 2011. "Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: A review." *Vision Research* 51 (8): 836–852.
- Sprague, W. W., E. A. Cooper, S. Reissier, B. Yellapragada, and M. S. Banks. 2016. "The natural statistics of blur." *Journal of Vision* 16 (10): 23.

- Stein, H. A. 1990. “The management of presbyopia with contact lenses: a review.” *Eye & Contact Lens* 16 (1): 33–42.
- Steinbach, M. J., and R. Held. 1968. “Eye Tracking of Observer-Generated Target Movements.” *Science* 161 (3837): 187–188.
- Steinicke, F., and G. Bruder. 2014. “A self-experimentation report about long-term use of fully-immersive technology.” In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction*, 66–69. ACM.
- Stewart, C., M. Moseley, A. Fielder, and D. Stephens. 2004. “Refractive adaptation in amblyopia: quantification of effect and implications for practice.” *British Journal of Ophthalmology* 88 (12): 1552–1556.
- Stone, R., and D. Flitcroft. 2004. “Ocular shape and myopia.” *Annals-Academy of Medicine Singapore* 33 (1): 7–15.
- Strang, N. C., M. Day, L. S. Gray, and D. Seidel. 2011. “Accommodation steps, target spatial frequency and refractive error.” *Ophthalmic and Physiological Optics* 31 (5): 444–455.
- Strang, N. C., B. Winn, and A. Bradley. 1998. “The role of neural and optical factors in limiting visual resolution in myopia.” *Vision Research* 38 (11): 1713–1721.
- Sugiura, A., K. Tanaka, H. Takada, T. Kojima, T. Yamakawa, and M. Miyao. 2015. “A Temporal Analysis of Body Sway Caused by Self-Motion During Stereoscopic Viewing.” In *International Conference on Universal Access in Human-Computer Interaction*, 246–254. Springer.

- Sweeney, J., M. Mintun, S. Kwee, M. Wiseman, D. Brown, D. Rosenberg, and J. Carl. 1996. "Positron emission tomography study of voluntary saccadic eye movements and spatial working memory." *Journal of Neurophysiology* 75 (1): 454–468.
- Taylor, V., M. Bossi, C. Bunce, J. A. Greenwood, and A. Dahlmann-Noor. 2015. "Binocular versus standard occlusion or blurring treatment for unilateral amblyopia in children aged three to eight years." *The Cochrane Library*.
- Taylor, V., M. Bossi, J. A. Greenwood, and A. Dahlmann-Noor. 2016. "Childhood amblyopia: current management and new trends." *British Medical Bulletin*.
- Takagi, M., H. Oyamada, H. Abe, D. S. Zee, H. Hasebe, A. Miki, T. Usui, S. Hasegawa, and T. Bando. 2001. "Adaptive changes in dynamic properties of human disparity-induced vergence." *Investigative Ophthalmology & Visual Science* 42 (7): 1479–1486.
- Tanaka, K., and H. Saito. 1989. "Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey." *Journal of Neurophysiology* 62 (3): 626–641.
- Taylor, C. P., and P. J. Bex. 2015. "On the number of perceivable blur levels in naturalistic images." *Vision Research* 115:142–150.

- Taylor, J., W. N. Charman, C. O'Donnell, and H. Radhakrishnan. 2009. "Effect of target spatial frequency on accommodative response in myopes and emmetropes." *Journal of Vision* 9 (1): 16–16.
- Thibos, L., D. Walsh, and F. Cheney. 1987. "Vision beyond the resolution limit: aliasing in the periphery." *Vision Research* 27 (12): 2193–2197.
- To, L., B. Thompson, J. R. Blum, G. Maehara, R. F. Hess, and J. R. Cooperstock. 2011. "A game platform for treatment of amblyopia." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19 (3): 280–289.
- Traver, V., and F. Pla. 2008. "Log-polar mapping template design: From task-level requirements to geometry parameters." *Image Vision Computing* 26 (10): 1354–1370.
- Troilo, D., and S. J. Judge. 1993. "Ocular development and visual deprivation myopia in the common marmoset (*Callithrix jacchus*)." *Vision Research* 33 (10): 1311–1324.
- Troilo, D., N. Quinn, and K. Baker. 2007. "Accommodation and induced myopia in marmosets." *Vision Research* 47 (9): 1228–1244.
- Troilo, D., K. Totonelly, and E. Harb. 2009. "Accommodation and the Visual Regulation of Refractive State in Marmosets." *Optometry & Vision Science* 86 (1): E31.
- Tsirlin, I., R. S. Allison, and L. M. Wilcox. 2008. "Stereoscopic transparency: Constraints on the perception of multiple surfaces." *Journal of Vision* 8 (5): 5–5.

- Ukai, K., and P. A. Howarth. 2008. "Visual fatigue caused by viewing stereoscopic motion images: Background, theories, and observations." *Displays* 29 (2): 106–116.
- Ukai, K., and A. Kibe. 2003. "Counterroll torsional eye movement in users of head-mounted displays." *Displays* 24 (2): 59–63.
- Ungerleider, L. G., and L. Pessoa. 2008. "What and where pathways." *Scholarpedia* 3 (11): 5342.
- Van Horn, M. R., P. A. Sylvestre, and K. E. Cullen. 2008. "The brain stem saccadic burst generator encodes gaze in three-dimensional space." *Journal of Neurophysiology* 99 (5): 2602–2616.
- Vedamurthy, I., D. C. Knill, S. J. Huang, A. Yung, J. Ding, O.-S. Kwon, D. Bavelier, and D. M. Levi. 2016. "Recovering stereo vision by squashing virtual bugs in a virtual reality environment." *Philosophical Transactions of the Royal Society B* 371 (1697): 20150264.
- Vera-Diaz, F. A., J. Gwiazda, F. Thorn, and R. Held. 2004. "Increased accommodation following adaptation to image blur in myopes." *Journal of Vision* 4 (12): 10–10.
- Vera-Diaz, F. A., G. Maiello, K. L. Kerber, F. Thorn, and P. J. Bex. 2015. "Myopes' ability to discriminate and detect blur." *Investigative Ophthalmology & Visual Science* 56 (7): 537–537.

- Vercher, J.-L., G. M. Gauthier, O. Guedon, J. Blouin, J. Cole, and Y. Lamarre. 1996. "Self-moved target eye tracking in control and deafferented subjects: roles of arm motor command and proprioception in arm-eye coordination." *Journal of Neurophysiology* 76 (2): 1133–1144.
- Vercher, J.-L., and G. Gauthier. 1992. "Oculo-manual coordination control: ocular and manual tracking of visual targets with delayed visual feedback of the hand motion." *Experimental Brain Research* 90 (3): 599–609.
- Vercher, J.-L., D. Quaccia, and G. Gauthier. 1995. "Oculo-manual coordination control: respective role of visual and non-visual information in ocular tracking of self-moved targets." *Experimental Brain Research* 103 (2): 311–322.
- Vergheze, P., T. L. Tyson, S. Ghahghaei, and D. C. Fletcher. 2016. "Depth Perception and Grasp in Central Field Loss." *Investigative Ophthalmology & Visual Science* 57 (3): 1476.
- Vienne, C., L. Sorin, L. Blondé, Q. Huynh-Thu, and P. Mamassian. 2014. "Effect of the accommodation-vergence conflict on vergence eye movements." *Vision Research* 100:124–133.
- Viirre, E., W. Cadera, and T. Vilis. 1988. "Monocular adaptation of the saccadic system and vestibulo-ocular reflex." *Investigative Ophthalmology & Visual Science* 29 (8): 1339–1347.
- Vishwanath, D. 2012. "The utility of defocus blur in binocular depth perception." *i-Perception* 3 (8): 541.



- Vishwanath, D., and E. Blaser. 2010. "Retinal blur and the perception of egocentric distance." *Journal of Vision* 10 (10): 26.
- Vitale, S., R. D. Sperduto, and F. L. Ferris 3rd. 2009. "Increased prevalence of myopia in the United States between 1971-1972 and 1999-2004." *Archives of Ophthalmology* 127 (12): 1632–1639.
- Volkman, F. C. 1962. "Vision during voluntary saccadic eye movements." *Journal of the Optical Society of America A* 52 (5): 571–577.
- Volkman, F. C., L. A. Riggs, K. D. White, and R. K. Moore. 1978. "Contrast sensitivity during saccadic eye movements." *Vision Research* 18 (9): 1193–1199.
- Volkman, F. C., A. M. Schick, and L. A. Riggs. 1968. "Time course of visual inhibition during voluntary saccades." *Journal of the Optical Society of America* 58 (4): 562–569.
- Von Helmholtz, H. 1867. *Handbuch der physiologischen Optik*. Voss.
- Vul, E., J. Bergsma, and D. I. A. MacLeod. 2010. "Functional adaptive sequential testing." *Seeing and Perceiving* 23 (5): 483–515.
- Walline, J. J., K. Lindsley, S. S. Vedula, S. A. Cotter, D. O. Mutti, and J. D. Twelker. 2011. "Interventions to slow progression of myopia in children." *The Cochrane Library*.
- Wallis, T. S. A., and P. J. Bex. 2012. "Image correlates of crowding in natural scenes." *Journal of Vision* 12 (7): 6.

- Wallman, J., M. Gottlieb, V. Rajaram, and L. Fugate-Wentzek. 1987. "Local retinal regions control local eye growth and myopia." *Science* 237 (4810): 73–77.
- Wallman, J., J. Turkel, and J. Trachtman. 1978. "Extreme myopia produced by modest change in early visual experience." *Science* 201 (4362): 1249–1251.
- Wallman, J., and J. Winawer. 2004. "Homeostasis of eye growth and the question of myopia." *Neuron* 43 (4): 447–468.
- Walsh, G., and W. Charman. 1988. "Visual sensitivity to temporal change in focus and its relevance to the accommodation response." *Vision Research* 28 (11): 1207–1221.
- Wang, B., and K. J. Ciuffreda. 2004. "Depth-of-focus of the human eye in the near retinal periphery." *Vision Research* 44 (11): 1115–1125.
- . 2005a. "Blur discrimination of the human eye in the near retinal periphery." *Optometry & Vision Science* 82 (1): 52–58.
- . 2005b. "Foveal blur discrimination of the human eye." *Ophthalmic and Physiological Optics* 25 (1): 45–51.
- . 2006. "Depth-of-focus of the human eye: theory and clinical implications." *Survey of Ophthalmology* 51 (1): 75–85.
- Wang, B., K. J. Ciuffreda, and T. Irish. 2006. "Equiblur zones at the fovea and near retinal periphery." *Vision Research* 46 (21): 3690–3698.

- Wang, J., M. Barkowsky, V. Ricordel, and P. Le Callet. 2011. “Quantifying how the combination of blur and disparity affects the perceived depth.” In *IS&T/SPIE Electronic Imaging*, 78650K–78650K. International Society for Optics and Photonics.
- Wann, J. P., S. Rushton, and M. Mon-Williams. 1995. “Natural problems for stereoscopic depth perception in virtual environments.” *Vision Research* 35 (19): 2731–2736.
- Warren Jr, W. H., and D. J. Hannon. 1988. “Direction of self-motion is perceived from optical flow.” *Nature* 336:162–163.
- Warren Jr, W. H., A. Blackwell, K. Kurtz, N. G. Hatsopoulos, and M. Kalish. 1991. “On the sufficiency of the velocity field for perception of heading.” *Biological Cybernetics* 65 (5): 311–320.
- Warren Jr, W. H., B. A. Kay, W. D. Zosh, A. P. Duchon, and S. Sahuc. 2001. “Optic flow is used to control human walking.” *Nature Neuroscience* 4 (2): 213–216.
- Warren Jr, W. H., M. W. Morris, and M. Kalish. 1988. “Perception of translational heading from optical flow.” *Journal of Experimental Psychology: Human Perception and Performance* 14 (4): 646.
- Warren Jr, W. H., and J. A. Saunders. 1995. “Perceiving heading in the presence of moving objects.” *Perception* 24 (3): 315–331.
- Warren, P. A., and S. K. Rushton. 2007. “Perception of object trajectory: parsing retinal motion into self and object movement components.” *Journal of Vision* 7 (11): 2–2.

- Warren, P. A., and S. K. Rushton. 2009. "Optic flow processing for the assessment of object movement during ego movement." *Current Biology* 19 (18): 1555–1560.
- Warren, W. H., and D. J. Hannon. 1990. "Eye movements and optical flow." *Journal of the Optical Society of America A* 7 (1): 160–169.
- Wasserstein, R. L., and N. A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70 (2): 129–133.
- Waterworth, J., E. L. Waterworth, G. Riva, and F. Manovani. 2010. "On feeling (the) present: an evolutionary account of the sense of presence in physical and electronically-mediated environments." *Journal of Consciousness Studies* 17 (1-2): 167–188.
- Watson, A. B., A. J. Ahumada Jr, et al. 1985. "Model of human visual-motion sensing." *Journal of the Optical Society of America A* 2 (2): 322–341.
- Watson, A. B., and A. J. Ahumada. 2011. "Blur clarified: a review and synthesis of blur discrimination." *Journal of Vision* 11 (5): 10.
- Watson, A. B., and J. I. Yellott. 2012. "A unified formula for light-adapted pupil size." *Journal of Vision* 12 (10): 12.
- Watt, R., and M. Morgan. 1983. "The recognition and representation of edge blur: evidence for spatial primitives in human vision." *Vision Research* 23 (12): 1465–1477.
- . 1984. "Spatial filters and the localization of luminance changes in human vision." *Vision Research* 24 (10): 1387–1397.

- Watt, S. J., K. Akeley, M. O. Ernst, and M. S. Banks. 2005. "Focus cues affect perceived depth." *Journal of Vision* 5 (10): 7–7.
- Watt, S. J., K. Akeley, A. R. Girshick, and M. S. Banks. 2005. "Achieving near-correct focus cues in a 3d display using multiple image planes." In *Electronic Imaging 2005*, 393–401. International Society for Optics and Photonics.
- Weichert, F., D. Bachmann, B. Rudak, and D. Fisseler. 2013. "Analysis of the accuracy and robustness of the leap motion controller." *Sensors* 13 (5): 6380–6393.
- Weissman, M. A., and A. J. Woods. 2011. "A simple method for measuring crosstalk in stereoscopic displays." *IS&T/SPIE Electronic Imaging* 7863:786310.
- Wetherill, G. B., and H. Levitt. 1965. "Sequential estimation of points on a psychometric function." *British Journal of Mathematical and Statistical Psychology* 18 (1): 1–10.
- Weymouth, F. W. 1958. "Visual sensory units and the minimal angle of resolution." *American Journal of Ophthalmology* 46 (1): 102–113.
- Wickens, T. D. 2002. *Elementary Signal Detection Theory*. Oxford University Press Oxford, UK.
- Wiesel, T. N., and E. Raviola. 1977. "Myopia and eye enlargement after neonatal lid fusion in monkeys." *Science* 266 (5597): 66–68.

- Wildsoet, C. F., H. C. Howland, S. Falconer, and K. Dick. 1993. "Chromatic aberration and accommodation: their role in emmetropization in the chick." *Vision Research* 33 (12): 1593–1603.
- Wildsoet, C., and J. Wallman. 1995. "Choroidal and scleral mechanisms of compensation for spectacle lenses in chicks." *Vision Research* 35 (9): 1175–1194.
- Wilkinson, M. O., R. S. Anderson, A. Bradley, and L. N. Thibos. 2016. "Neural bandwidth of veridical perception across the visual field." *Journal of Vision* 16 (2): 1–1.
- Wood, J. M., M. J. Collins, A. Chaparro, R. Marszalek, T. Carberry, P. Lacherez, and B. S. Chu. 2014. "Differential Effects of Refractive Blur on Day and Nighttime Driving Performance." *Investigative Ophthalmology & Visual Science* 55 (4): 2284–2289.
- Wright, K. W., P. H. Spiegel, and L. S. Thompson. 2006. *Handbook of pediatric strabismus and amblyopia*. Springer.
- Wright, M., and A. Johnston. 1983. "Spatiotemporal contrast sensitivity and visual field locus." *Vision Research* 23 (10): 983–989.
- Wurbs, J., E. Mingolla, and A. Yazdanbakhsh. 2013. "Modeling a space-variant cortical representation for apparent motion." *Journal of Vision* 13 (10): 2.
- Xiao, D. K., V. L. Marcara, and G. A. Rangel. 1997. "Sensitivity of Macaque MT/V5 Neurons for Surface Orientation in Depth Specified by Motion." *European Journal of Neuroscience* 9:956–964.

- Xu, H., P. Wallisch, and D. C. Bradley. 2014. "Spiral motion selective neurons in area MSTd contribute to judgments of heading." *Journal of Neurophysiology* 111 (11): 2332–2342.
- Xu, P., K. Ehinger, Y. Zhang, A. Finkelstein, S. Kulkarni, and J. Xiao. 2015. "TurkerGaze: Crowdsourcing Saliency with Webcam based Eye Tracking." *arXiv preprint arXiv:1504.06755*.
- Yamada, T., and K. Ukai. 1997. "Amount of defocus is not used as an error signal in the control system of accommodation dynamics." *Ophthalmic and Physiological Optics* 17 (1): 55–60.
- Yamaguchi, T., K. Ohnuma, K. Konomi, Y. Satake, J. Shimazaki, and K. Negishi. 2013. "Peripheral optical quality and myopia progression in children." *Graefe's Archive for Clinical and Experimental Ophthalmology* 251 (10): 2451–2461.
- Yang, Q., and Z. Kapoula. 2003. "Binocular coordination of saccades at far and at near in children and in adults." *Journal of Vision* 3 (8): 3.
- Yeh, Y.-Y., and L. D. Silverstein. 1990. "Limits of fusion and depth judgment in stereoscopic color displays." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 32 (1): 45–60.
- Young, M. J., M. S. Landy, and L. T. Maloney. 1993. "A perturbation analysis of depth perception from combinations of texture and motion cues." *Vision Research* 33 (18): 2685–2696.

- Young, S. D., B. D. Adelstein, and S. R. Ellis. 2007. “Demand characteristics in assessing motion sickness in a virtual environment: or does taking a motion sickness questionnaire make you sick?” *IEEE Transactions on Visualization and Computer Graphics* 13 (3): 422–428.
- Zhang, T., L. O’hare, P. B. Hibbard, H. T. Nefs, and I. Heynderickx. 2014. “Depth of field affects perceived depth in stereographs.” *ACM Transactions on Applied Perception (TAP)* 11 (4): 18.
- Zhou, W., and W. M. King. 1998. “Premotor commands encode monocular eye movements.” *Nature* 393 (6686): 692–695.



# Appendix

## List of Conference Posters and Presentations

- **Maiello, G.**, Chessa, M., Solari, F., & Bex, P. J. (2013) The Contribution of Perspective, Blur and Disparity to Depth Perception in Natural Vision. [Talk presentation] Association for Research in Vision and Ophthalmology, Seattle, Washington, USA - May 05 - 09, 2013
- **Maiello, G.**, Chessa, M., Solari, F., & Bex, P. J. (2013) Stereoscopic fusion with gaze-contingent blur. [Poster presentation] European Conference on Visual Perception, Bremen, Germany - August 25 - 29, 2013
- **Maiello, G.**, Harrison, W. J., & Bex, P. J. (2014) Dichoptic saccadic adaptation. [Poster presentation] Vision Sciences Society, St. Pete Beach, Florida, USA - May 16 - 21, 2014
- Harrison, W. J., **Maiello, G.**, & Bex, P. J. (2014). Illusory objects are altered by saccadic eye movement preparation. [Poster presentation] Neuroscience 2014, Washington, DC, USA - November 15 - 19, 2014
- Vera-Diaz, F. A., **Maiello, G.**, Kerber, K.L., Thorn, F., & Bex, P. J. (2015) Myopes' ability to discriminate and detect blur. [Poster presentation] Association for Research in Vision and Ophthalmology, Denver, Colorado, USA - May 03 - 07, 2015
- Kerber, K.L., **Maiello, G.**, Thorn, F., Bex, P. J., & Vera-Diaz, F. A. (2015) Myopes' ability to accurately accommodate to blur cues in virtual 3D images. [Poster presentation] Association for Research in Vision and Ophthalmology, Denver, Colorado, USA - May 03 - 07, 2015

- **Maiello, G.**, Chessa, M., Bex, P. J., & Solari, F. (2015) A Space-Variant Model for Motion Interpretation across the Visual Field. [Talk Presentation] MODVIS 2015 Computational and Mathematical Models in Vision, St. Pete Beach, Florida, USA - May 13 - 15, 2015
- **Maiello, G.**, Harrison, W. J., Vera-Diaz, F. A., & Bex, P. J. (2015) Perceptual Consequences of Elongated Eyes. [Poster presentation] Vision Sciences Society, St. Pete Beach, Florida, USA - May 15 - 20, 2015
- Walker, L., **Maiello, G.**, Bex, P. J., & Vera-Diaz, F. A. (2015) Ability to perceive blur in central and near-peripheral retina in myopia. [Poster presentation] Association for Research in Vision and Ophthalmology, Seattle, Washington, USA - May 01 - 05, 2016
- Gibaldi, A., **Maiello, G.**, Bex, P. J., & Sabatini, S. P. (2016) Modelling Short-Latency Disparity-Vergence Eye Movements Under Dichoptic Unbalanced Stimulation. [Talk Presentation] MODVIS 2016 Computational and Mathematical Models in Vision, St. Pete Beach, Florida, USA - May 11 - 13, 2016
- Carroll, S., **Maiello, G.**, Harrison, W. J., & Bex, P. J. (2016) Myopic Eyes See Better in a Crowd. [Poster presentation] Vision Sciences Society, St. Pete Beach, Florida, USA - May 13 - 18, 2016
- **Maiello, G.**, Kwon, M., & Bex, P. J. (2016) 3 Dimensional Binocular Eye and Hand Coordination in Normal Vision and with Simulated Visual Impairments. [Talk presentation] Vision Sciences Society, St. Pete

Beach, Florida, USA - May 13 - 18, 2016

- **Maiello, G.**, Chessa, M., Bex, P. J., & Solari, F. (2016) Can Neuro-morphic Computer Vision Inform Vision Science? Disparity Estimation as a Case Study. [Talk Presentation] MODVIS 2016 Computational and Mathematical Models in Vision, St. Pete Beach, Florida, USA - May 11 - 13, 2016
- **Maiello, G.**, Gibaldi, A., Sabatini, S. P., & Bex, P. J. (2016) Vergence Eye Movements to Unbalanced Dichoptic Visual Stimuli. [Talk presentation] European Conference on Visual Perception, Barcelona, Spain - August 28 - September 1, 2016
- Walker, L., **Maiello, G.**, Bex, P. J., Coletta, N., & Vera-Diaz, F. A. (2016) Vergence Blur Perception in Near-Peripheral Retina. [Poster presentation] American Academy of Optometry, Anaheim, California, USA - November 9 - November 12, 2016

## List of Peer-Reviewed Publications

- **Maiello, G.**, Chessa, M., Solari, F., & Bex, P. J. (2014). Simulated disparity and peripheral blur interact during binocular fusion. *Journal of vision*, 14(8), 13. doi: 10.1167/14.8.13
- **Maiello, G.**, Chessa, M., Solari, F., & Bex, P. J. (2015). The (In)Effectiveness of Simulated Blur for Depth Perception in Naturalistic Images. *PloS one*, 10(10), e0140230. doi: 10.1371/journal.pone.0140230
- Gibaldi, A, Vanegas, M., Bex P. J., & **Maiello, G.** (2016). Evaluation of the Tobii EyeX Eye Tracking Controller and Matlab Toolkit for Research. *Behavior Research Methods* (1-24). doi: 10.3758/s13428-016-0762-9
- Chessa, M., **Maiello, G.**, Bex, P. J., & Solari, F. (2016). A space-variant model for motion interpretation across the visual field. *Journal of vision*, 16(2), 12. doi: 10.1167/16.2.12
- **Maiello, G.**, Harrison, W.J., & Bex, P. J. (2016). Monocular and Binocular Contributions to Oculomotor Plasticity. *Scientific Reports* 6, 31861. doi: 10.1038/srep31861
- Chessa, M.†, **Maiello, G.**†, Borsari, A., & Bex P. J. (2016). The perceptual quality of the Oculus Rift for immersive virtual reality. *Human Computer Interaction* (in press). doi: 10.1080/07370024.2016.1243478

†co-first authors